Foreign and Domestic Private Investments in Developing Economies

Cases in sub-Saharan Africa

By

Paul Wabiga

Dissertation presented for the award of the degree of Doctor of Philosophy (Economics) in the Faculty of Economic and Management Sciences at Stellenbosch University

Verhandeling aangebied vir die toekenning van die graad Doktor in Filosofie (Ekonomiese) aan die Fakulteit Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

Supervisor: Professor Neil Rankin

March 2020

Declaration

By electronically submitting this thesis, I declare that all the work contained herein is my own original work, that I am the original owner thereof (unless to the extent clearly otherwise stated) and that I have not previously in its entirety or in part submitted it to any other institution for any award whatsoever.

Name: Paul Wabiga

Date: March 2020

Dedication

To my family. My wife, Dianah Birungi; my children; Noel, Noelia, Noelina, and Noeline (RIP), I will forever be grateful for your support and patience during my rather difficult study period.

Acknowledgements

Without the immense support, mentoring, encouragement, and guidance from my supervisor, Professor Neil Rankin, the work in this thesis would have been impossible. Throughout the programme years, Neil has tirelessly guided me and imparted knowledge unto me, making the development of this noble work possible. His support and encouragement during the often-challenging doctoral process formed the pivot upon which my success rotated. With Neil choosing to work with and guide me, I got the opportunity of admission to Stellenbosch University's Graduate School of Economic and Management Sciences (GEM), along with a three-year scholarship for doctoral studies at this very important level. I intend to uphold what he has passed unto me in terms of knowledge, interpersonal skills and work ethics in my future career as an economist.

I am grateful to my fellow students in the 2017 GEM PhD cohort; Leon Matsuro, Molise Thabang, Helvi P. Ndilimeke, Neema Robert, Nel Hienrich, Rossouw van Jaarsveld, Joel Gondwe, and Hassan Kimbugwe. Along with your crucial emotional support and encouragement, you provided wonderful debates and advice on economic issues pertaining to this work during weekly presentations. I am blessed to have met all of you in my journey of academic pursuit. Also, great appreciation goes to the World Bank and the Centre for the Study of African Economies (CSAE) at Oxford University for the data used in this thesis. I am equally grateful to Dr Adam Rauh of Massachusetts Institute of Technology for your guidance in Panelmatch applications using R.

Special thanks to Dr John Sseruyange, Prof. Eria Hisali, Dr Faisal Buyinza, Dr Wasswa Francis, Prof. Edward Bbaale, and all my other friends and colleagues, who provided valuable insights and beneficial support before and during my studies. I particularly wish to single out Dr Sseruyange, Dr Sarah Nakijoba, and Dr Jude Mugarura, to whom I am grateful for the time spent deliberating on numerous economic issues, PhD distresses and life lessons. I hope to take forward both my professional and personal relationships that have been cemented during these past years of my study. I am also, in a special way, indebted to Dr Fred K. Muhumuza, Dr Madinah M. Guloba (EPRC), Dr Tonny Muhumuza (UNDP-Uganda) and Mr Stefan Kratzsch (UNIDO) for your enthusiastic assistance during my search for firm-level data.

It should be noted that opinions expressed and conclusions arrived at in this thesis are those of the author, and should not to be attributed to GEM, CSAE or the World Bank. I am appreciative for the financial support of the GEM full-time doctoral scholarship, without which my studies would have been interminably tougher to undertake. In this regard, I extend my special appreciation to the GEM Manager, Dr Jaco Franken, for his immensely supportive role during the three years of my studies. You are wonderful. Be blessed.

Last, but not least, my golden family has provided years of support and reassurance. My wife, children, parents, brothers, sisters and in-laws have been pillars of strength – financially, physically and emotionally – throughout the process. I am grateful to have been treasured so completely by such wonderful people. To my daughter, Noeline, I thank the Lord for the time he allowed us share moments with you, and pray that your soul rests in peace. My friends have been of great help, providing both motivation and much-needed diversions, bringing me back down to earth when I became overwhelmed. My parents-in-law have provided additional support in many ways, for which I am endlessly grateful. I would like to thank my wife specifically, for managing with my long absence from home during my three-year study period and the development of this thesis; and especially for managing our family, giving life to our children and to me. Her belief and pride in my abilities have kept me enthused and encouraged me to do the best that I could for our future.

Finally, I would like to thank and give praise, honour and glory to the Almighty God for His providence. All the above-mentioned contributions would have yielded nothing without His will.

Abstract

The role of foreign-owned investments in the economic transformation of host economies cannot be overstated. Foreign-owned investments are associated with higher levels of employment; higher mean wages; technological transfer via spill-overs to domestic firms; high capital and export intensities. Through these positive features, foreign investments are generally associated with positive effects on economic growth. Economies in Sub-Saharan Africa (SSA) have been cognisant of these positive benefits, as is manifested by their national development plans for the near future. As more foreign-owned investments have flowed in the region, there has been visible economic growth, but with dismal effects in terms of employment creation in both numbers and quality. Exportation remains low and the impact on poverty remains mild. These poor results undermine the strength of the supposedly empirical association between foreign-owned investments and economic transformation of typical host economies, at least in SSA.

Existing empirical knowledge on the nature of foreign-owned investments in SSA and their likely effects on economic welfare of host economies has either remained scanty or mainly been at the macro level due to limited data. This thesis attempts to sidestep this challenge by utilizing three different firm-level data sets to examine the nature and effects of foreign investments in relation to domestic firms in SSA from a microeconomic perspective. Firstly, a relatively novel, unsupervised, machine-learning approach has been used to classify firms so as to study their characteristic features. In this attempt, firm-level performance features, which have empirically been found by existing studies to distinguish foreign from domestic firms, are re-examined. This inquiry is a robustness check for previous studies and is undertaken from both a one-country and a multi-country perspective. Most studies that have utilised firm-level data to classify firms have in most cases been based on assumptions specified *a priori*. Using performance indicators such as output, employment and exports, correlation results from these studies are reported based on these classifications. In this thesis, firms are classified and their performance features examined without prior assumptions set especially regarding data distribution. Agglomerative clustering methods have been used to generate groups of firms *a posteriori* before examining these groups descriptively along performance indicators. Key findings indicate that foreign-owned firms systematically differ from domestically owned firms along numerous performance indicators, while there is a high likelihood of intra-foreign-owned firms' heterogeneities.

In the second and final analysis, this thesis has employed regression and matching techniques with difference-in-differences estimation methods to investigate the existence and nature of effect of foreign ownership on firm-level performance in SSA. This thesis considers foreign ownership arising out of acquisition of a local firm by foreigners. This provides crucial evidence as to whether foreign-owned investments in SSA truly influence host economies' welfare from a firm-level perspective. Empirical findings in this thesis indicate that, although cream-skimming is prevalent in foreign acquisition of formerly domestically owned firms, foreign ownership has positive effects on firm performance. This thesis finds positive acquisition effects on employment, wages, productivity, output, skill intensity, and capital investments. These positive effects are significant for wages, output, and productivity. By implication, through their effects at firm-level, foreign-owned investments are still a potential channel through which sustained welfare enhancements in SSA can be achieved, given well-intentioned policies being in place.

Abstrakte

Die rol van beleggings in buitelandse besit in die ekonomiese transformasie van gasheerekonomieë kan nie oorbeklemtoon word nie. Beleggings in buitelandse besit hou verband met hoër indiensnemingsvlakke; hoër gemiddelde lone; tegnologiese oordrag via stortings na huishoudelike firmas; hoë kapitaal- en uitvoerintensiteite. Deur hierdie positiewe eienskappe word buitelandse beleggings meestal geassosieer met 'n positiewe uitwerking op ekonomiese groei. Hierdie positiewe voordele is bewus van die ekonomieë in Afrika suid van die Sahara (SSA), soos blyk uit hul nasionale ontwikkelingsplanne vir die nabye toekoms. Namate meer beleggings in buitelandse besit in die streek gevloei het, was daar sigbare ekonomiese groei, maar met 'n aaklige uitwerking wat betref die skepping van indiensneming in getalle en kwaliteit. Uitvoer bly laag en die impak op armoede bly mild. Hierdie swak resultate ondermyn die sterkte van die vermeende empiriese verband tussen beleggings in buitelandse besit en ekonomiese transformasie van tipiese gasheerekonomieë, ten minste in SSA.

Bestaande empiriese kennis oor die aard van buitelandse beleggings in SSA en die waarskynlike gevolge daarvan vir die ekonomiese welvaart van gasheerekonomieë het nogal gebly of hoofsaaklik op makro-vlak gebly as gevolg van beperkte gegewens. Hierdie tesis poog om hierdie uitdaging te benadeel deur drie verskillende datastelle te gebruik om die aard en gevolge van buitelandse beleggings met betrekking tot plaaslike ondernemings in SSA vanuit 'n mikro-ekonomiese perspektief te ondersoek. Eerstens is 'n relatief nuwe, sonder toesig, masjienleer-benadering gebruik om ondernemings te klassifiseer om hul kenmerkende kenmerke te bestudeer. In hierdie poging word prestasie-funksies op firma-vlak, wat deur bestaande studies empiries gevind is om buitelandse en plaaslike ondernemings te onderskei, heroorweeg. Hierdie ondersoek is 'n robuustheidsondersoek vir vorige studies en word onderneem vanuit 'n een-land- en 'n multi-landse perspektief. Die meeste studies wat data op firma-vlak gebruik het om ondernemings te klassifiseer, is in die meeste gevalle gebaseer op a priori-aannames. Op grond van hierdie klassifikasies word prestasie-aanwysers soos uitset, indiensneming en uitvoere gerapporteer. In hierdie tesis word firmas geklassifiseer en hul prestasie-eienskappe ondersoek sonder voorafgaande aannames, veral met betrekking tot dataverspreiding. Agglomeratiewe groeperingsmetodes is gebruik om groepe firmas a posteriori te genereer voordat hierdie groepe beskrywend met prestasie-aanwysers ondersoek word. Belangrike bevindings dui aan dat ondernemings in buitelandse besit stelselmatig verskil

van ondernemings in plaaslike besit volgens talle prestasie-aanwysers, terwyl die groot waarskynlikheid bestaan dat ondernemings in buitelandse besit se heterogeniteit bestaan.

In die tweede en finale ontleding het hierdie tesis regressie- en bypassingstegnieke gebruik met beramingmetodes van difference-in-differences om die bestaan en aard van die effek van buitelandse eienaarskap op firma-vlakprestasie in SSA te ondersoek. In hierdie proefskrif word buitelandse eienaarskap voortspruitend uit die verkryging van 'n plaaslike firma deur buitelanders. Dit lewer 'n belangrike bewys of beleggings in buitelandse besit in SSA die welvaart van die gasheerekonomie werklik vanuit 'n vaste perspektief beïnvloed. Empiriese bevindings in hierdie proefskrif dui aan dat, hoewel kremskeep algemeen voorkom in die buitelandse verkryging van voorheen plaaslike ondernemings, buitelandse eienaarskap 'n positiewe uitwerking op die onderneming se prestasie het. Hierdie tesis vind verkrygingseffekte op indiensneming, lone, produktiwiteit, produksie en kapitaalbeleggings. Hierdie positiewe gevolge is beduidend vir lone, produksie en produktiwiteit. By implikasie is beleggings in buitelandse besit deur die gevolge daarvan op firma-vlak steeds 'n potensiële kanaal waardeur volgehoue welsynverbeterings in SSA bereik kan word, gegewe die goedbedoelde beleid wat bestaan.

Table of Contents

Contents

List of Tables

List of Figures

List of Appendix Tables and Figures

Appendix Figures

Chapter 1

Introduction

1.1 Motivation and Context

Private investments, both foreign-owned and domestically-owned, have been a subject of empirical investigation for a long time and literature on these investments continues to grow. Foreign-owned firms, which are popularly referred to as foreign direct investments (FDIs) in literature have increasingly been studied in comparison with domestically-owned investments in typical host-economy environments. Many of the empirical studies have focused on, among other aspects, firm-level characteristics and performance of these investments. Specific attention has, among other aspects, been on characteristic features of foreign-owned investments relative to those that are domestically-owned. Attention has also been on the firm-level performance of these investments, specifically the superior performance of foreign-owned investments when compared to domestically-owned plants. Some studies have also focused on whether and how foreign-owned investments affect performance of domestically-owned investments particularly via spill-overs. Empirical investigation on these investments has been recently promoted mainly by increased access to firm-level data, especially in the developed world.

Historically, however, empirical literature on foreign-owned investments or FDIs, particularly on their characteristics and performance in relation to domestically-owned investments, has its roots in the seminal work of Stephen Hymer in the 1960s. Hymer's work (Hymer, 1960) was the first to highlight the typical differences between foreign-owned and domestically-owned firms. He shed light on how these differences, manifesting mainly in terms of firms' structural features, were the most likely explanation for foreign-owned investments' usually superior performance. Along with other aspects, Hymer highlighted the fact that foreign-owned firms have specific ownership advantages that tend to enable them to perform relatively better than domestically-owned firms in a typical host economy environment. Hymer's work led to the emergence of more scholarly investigations into the structural and performance features of FDIs, including Kindleberger (1969), Aliber (1970), Caves (1971), Knickerbocker (1973), Rugman (1981), Casson (1983) and Dunning (1993).

Occurring simultaneously with access to firm-level data was the development of further theoretical formulations. These theories attempted to explain behavioural characteristics of foreign-owned firms in relation to domestically-owned firms. A case in point is Buckley and Casson's (1976) internalisation theoretical model, which is based on earlier works by Coase (Coase, 1937). John H Dunning (1988) later developed the eclectic theory, building on the strengths of the internalization theory and other previous theories. He incorporated time, location and ownership dimensions in his new framework. The eclectic theory is now the most popularly used in analyses.

With access to data and the existence of theoretical formulations, it is not surprising that literature on foreign-owned and domestically-owned private investments has continued to grow. As a result, it is now a stylized fact that foreign-owned investments tend to differ from and are likely to perform relatively better than domestically-owned firms on several performance indicators (Peluffo, 2015: 1). Findings in literature associate foreign-owned investments with; higher wages, higher productivity levels, a tendency to being more export oriented, employment of more workers, and being more skill and capital intensive when compared to domestically-owned firms. Studies by Almeida (2007), Girma and Gorg (2007a), Matthias and Javorcik (2009) and Njikam (2018) have confirmed these firm-level performance features among foreign-owned firms. However, these studies, along with many more that have examined these investments, do not clarify whether the characteristic and performance differences between foreign-owned and domestically-owned firms are a consequence of foreign ownership; nor do they consider whether, especially for non-greenfield investments, foreign investors, during acquisition of previously domestically-owned firms, merely target firms that are already performing strongly, resulting in post-acquisition differences that may not be due to foreign ownership itself. Additionally, these studies do not clarify whether the observed characteristic differences between the two types of firms are simply a reflection of comparisons between sub-groups of firms (such as between firms in high wage sectors, where foreign-owned firms are likely to be concentrated, and firms in lower wage sectors) or systematic differences exist between the two types of investment firms. Moreover, if foreign ownership supposedly accounts for the observed differences and consequent performance outcomes – this being the causal effect – its magnitude is still inadequately delineated, especially in sub-Saharan Africa (SSA). In delineating such effects, it is prudent to note that it is not easy to rule out the influence of 'cream-skimming' tendencies on the acquisition decision if foreign-ownership arises out of acquisition of a formerly domestically-owned firm in a typical host economy. Should such 'cream-skimming' be the case, then any attempt to ascribe the resulting performance to foreign-ownership itself will be subject to selection bias.

Scholars have attempted to answer some of the questions outlined above but most analysts have focused mainly on the developed, Asian and Latin American economies. This may be because most foreign-owned investments across the globe are, among other reasons, located in these economies. Few studies have focused on low developing countries, especially SSA. Data availability has been cited as one of the hindrances to the growth of empirical literature on foreign-owned firms in relation to domestically-owned firms in SSA. Nevertheless, empirical findings from the few studies conducted in SSA tend, to some extent, mirror those addressing other parts of the world. The findings of empirical investigations conducted in SSA by Rankin, Söderbom and Teal (2005), Coniglio, Prota and Seric (2015), Blanas, Seric and Viegelahn (2017), and Njikam (2018) echo those of the developed world, along with some emerging economies.

In the last decade, SSA has witnessed increased inflows of foreign investments. However, despite such increased inflows of late, anticipated benefits as alluded to in the previous discussion like export volumes remain dismal, being mainly in primary commodities (International Monetary Fund, 2018: 15), and with manufactured goods only accounting for 19% of Africa's exports to the rest of the world in 2015 (International Monetary Fund, 2018: 23). This contradicts for example the belief that economies with foreign investments are more likely to export finished products to the world market (Libanda, Nyasa & Marshall, 2017: 4). In 2018, labour productivity growth in SSA remained at 0.6%, which is below the average of 3.1% in the rest of the world; and employment growth in SSA only averaged at 3.1% between 2017 and 2019 (International Labor Organization, 2019: 28). This implies that the expected benefits/features associated with foreign-owned investments in relation to domestically-owned investments remain unimpressive even as more foreign investments continue to flow into the region. The unimpressive outcomes amidst increased inflows of foreign investments, together with the questions alluded to in the previous discussion, re-ignite the debate about the likely effects of foreign ownership on firm-level performance and hence justify further empirical examination of private investments at least in SSA. Are the foreign-owned investments flowing into the region characteristically superior to domestically-owned investments as empirical literature suggests for other parts of the world? If this is the case, then as more investments flow into SSA, the expectation would be corresponding outcomes like more employment, increased exports, wages, among others. Such outcomes can either be directly from newly set-up investments or after partial or full takeover of a hitherto domestically-owned investments whose performance would be expected to improve. This thesis attempts to further empirical examination and contribution to the growing literature on foreign-owned and domestically-owned private investments in developing economies, focusing on SSA where, as evidence shows, economic outcomes seem not corresponding to increased inflows of foreign investments as available empirical studies tend to predict. The thesis furthers the empirical literature through micro analyses of firms since performance of economies is simply a reflection of behaviour and hence performance of micro-production units like firms. This thesis therefore utilises three unique but comparable firm-level data sets from; the World Bank, United Nations Industrial Development Organisation (UNIDO), and the Centre for the Study of African Economies (CSAE) at Oxford University. The World Bank data set is a comprehensive survey of firms within Uganda's manufacturing and service sectors. The UNIDO data set is a comprehensive survey of over 6000 firms from 19 countries in SSA, providing an opportunity to investigate phenomena from a multi-country context. The third data set is a panel data set on manufacturing firms in Ghana, collected in six rounds by the World Bank and the Oxford CSAE under the Regional Program for Enterprise Development (RPED) project.

Using the three unique but comparable data sets, this thesis provides evidence on two key empirical questions, which are; (i) whether, as available empirical studies have confirmed elsewhere, the two types of private investments systematically differ from each other in terms of firm-level performance and characteristic features in SSA economies and (ii) whether foreign ownership has any effect on firm-level performance in SSA as it has been confirmed or dispelled elsewhere. The expectation was that answers to these questions would further empirical understanding of foreign-owned in relation to domestically-owned investments and provide critical insights into the noncorresponding (to increased FDI inflows) economic outcomes in SSA. Whether the deviation from what available literature says should be attributed to other factors or to the least, scholarly investigation should delineate the extent to which foreign ownership ought to be associated with performance of firms.

Guided by the available data sets, the thesis studies the first empirical question from two contexts, one being country (Uganda) specific and the other, a multi-country context. The second empirical question is also studied from a specific country (Ghana) context. The choice of the geographical scope (specific country/multi-country) in this thesis was motivated by mainly data availability and the need to focus on SSA. Ghana specifically, provided the opportunity of investigating the effect of foreign ownership on firm performance in a small developing economy. Ghana had also adopted a new policy that exogenously influenced hitherto domestically owned firms to change ownership either partially or wholly to foreign ownership, which enabled me to estimate empirically the effect of changeover from domestic to foreign ownership on firm-level performance.

The first empirical question is answered in both Chapter 2 and Chapter 3. In Chapter 2, the empirical research question is answered by re-examining the characteristic features of foreign-owned firms and investigating whether these firms systematically differ from domestically-owned firms along selected performance and characteristic features. This investigation confirms whether or not, comparing the two types of firms comes down to merely comparing sub-groups of firms as earlier hinted on. This analysis is based on a specific economy, being Uganda. The choice of Uganda as the case study was motivated by mainly data availability and the fact that the situational evidence alluded to earlier in SSA is comparable to that in Uganda as it will be seen later. Elsewhere, empirical investigation on foreign direct investments sometimes necessitates focus on specific countries and firm characteristics rather than a general orientation of firms and countries (O'Brien & Williams, 2013: 195). To achieve the main objective of this chapter, unlike earlier studies that have employed conventional econometric methods, in this thesis unsupervised machine learning techniques are used to identify homogeneous clusters of firms. These clusters have been analysed in terms of average performance across numerous performance indicators using descriptive statistics. The use of cluster analysis methodologies has gained ground in areas of psychology and clinical medicine (Everitt, Landau, Leese & Stahl, 2011) but has only been minimally utilized in the fields of business and economics. The expectation in this chapter was that such investigation would provide evidence to validate or dispel, earlier empirical findings, moreover after employing alternative methods of analysis that have been not widely utilised in economics. Indeed, after applying the preferred methods of analysis on Ugandan firms, I found results that are largely similar to those in the available international literature on foreign-owned firms in relation domestically-owned firms. Findings specifically indicate that foreign-owned firms do indeed differ from domestically-owned firms systematically along numerous performance and characteristic variables. The chapter finds no significant differences amongst domestically-owned firms themselves, except in terms of workers' education. In the segmentation process, foreign ownership is found to be a very important factor in firm classification, even when it is correlated with numerous variables. The findings from this chapter might be useful for policy makers in Uganda and SSA in general since these economies aim to attract increased foreign investment inflows. These results justify efforts aimed at attracting and incentivizing foreign investments in Uganda; and can probably be extended to apply to relatively comparable economies. A key contribution of Chapter 2 is the use of alternative methods of analysis; specifically, cluster analysis using machine learning. Unlike previous studies whose analyses are largely based on assumptions especially regarding data distribution, cluster analysis lets the data itself, irrespective of distribution, determine how firms are classified before resulting groups are analysed. As earlier noted, these methods have previously not been widely used in economic analyses and so, their use provides an effective robustness check on earlier empirical findings.

Chapter 3 replicates the methods used in Chapter 2 on a broader data set covering 19 countries in SSA. Because private investments, especially foreign direct investments are topical globally, findings of investigations on such firms may vary across relatively similar economies even when the same phenomenon is investigated. Additionally, country-specific investigations may be subject to small sample data problems, irrespective of the methods used in the analyses. For these reasons, Chapter 3 uses a broader data set to examine the same empirical question as in Chapter 2, but from a multi-country perspective. The specific aim is to investigate whether the systematic differences between foreign-owned firms and domestically-owned firms as revealed in Chapter 2 are or can be unique to firms in a specific country, in this case Uganda, or instead, a phenomenon that probably typifies firms in relatively comparable economies in SSA. In furtherance of scholarly understanding, I also aim to establish whether study findings in Chapter 2 hold for internationally known economic groupings of economies, such as classifications by the World Bank and International Monetary Fund. After employing methods similar to those in Chapter 2, results in Chapter 3 largely reflect those in Chapter 2. However, in addition to these findings, results in Chapter 3 reveal intra-foreign-owned firms' heterogeneities, a result not found in Chapter 2 probably due to the smaller data set used. Exportation is found as a key segmenting variable amongst foreign-owned firms and hugely explains these intra- foreign-owned firms' differences. This may point to critical differences among foreign-owned firms themselves probably on even other firm-level characteristics. This may also have sound implications for policy targeting, since even foreign-owned firms may not be homogenous along the superior performance and characteristic features associated with them in the literature. However, in terms of general findings, foreign-owned firms are found to be systematically different from domestically-owned firms even when country groupings are incorporated in the analysis, for instance, lower-income versus middle-income groups. This may provide support for policy replication across some economies in SSA. Though important, findings on the systematic differences between foreign-owned and domestically owned firms in Chapters 2 and 3 do not adequately provide evidence on the magnitude of the effect of foreign ownership on firm-level performance. This makes it hard to proportionately ascribe the resulting economic performance outcomes detailed earlier to foreign investments and/or other factors thought to be correlated with foreign ownership at firm level. This other key empirical gap/question is what Chapter 4 of this thesis attempts to close/answer.

Chapter 4 uses panel data to examine the effect of foreign ownership on performance of firms. I specifically consider foreign ownership arising from acquisition of previously domestically-owned firms. This analysis provides empirical insights into the direction and magnitude of the effect of foreign ownership which, to the best of my knowledge, is still inadequately examined in the developing world, especially in SSA. As with Chapter 2, Chapter 4 utilises firm-level data on a specific country, Ghana. According to Naughtin and Rankin (2016), these types of investigations are very important for economic policy makers but require comprehensive and representative micro data, an opportunity the Ghana case provided. Therefore, Chapter 4's contribution to this thesis is the use of a 12-year-long panel data set of Ghanaian manufacturing plants to empirically estimate the effect of foreign ownership on firm-level performance. It provides evidence on the extent to which firm-level performance should be ascribed to ownership status per se. Given the inherent sample size issues and the implications this might have on the results, after careful utilization of regression methods to provide the needed evidence, analysis in this chapter applies some of the popular traditional impact evaluation methods to provide robustness checks of the main results. Specifically, the analysis involves application of matching and difference-in-differences estimation methods for the purpose of robustness checks on main regression findings. These methods enabled me to handle possible endogeneity arising from selection bias pertaining to the acquisition decision, un-observed firm-level heterogeneity that might influence outcomes of interest and other econometric issues.

Empirical findings from chapter 4 indicate that the likelihood of 'cherry-picking' tendencies during foreign investors' acquisition of domestically-owned firms is high. In short, foreign investors tend to target high-performing firms during the acquisition process, a fact not adequately highlighted by earlier studies in SSA. Additionally, the chapter reveals positive and significant effects of foreign acquisition on the wages, productivity, output, and capital investments of acquired firms in Ghana. Effects on performance outcomes such as employment and skill intensity are positive although not statistically significant. A surprising result is on capital intensity, which totally contradicts predictions of firm production theory. This may be due to sample size issues and missing data problems. Another finding worth noting from the matching specifically is that using this estimation method, wage increases are not associated with increases in labour productivity, contrary to the predictions of mainstream firm theory. This might be attributed to the weakness and impreciseness of this method as noted in econometric literature. All in all, these results are largely in agreement with numerous scholarly investigations that have been conducted outside SSA. These findings further justify the need for economic policies aimed at attracting more foreign investment inflows to take advantage of the positive effects from these investments, especially in SSA. Most importantly, results tend to imply that the dismal economic performance outcomes detailed earlier and which motivated the rigorous empirical inquiry in this thesis, may not entirely be ascribed to increases in foreign investment inflows but also to numerous other factors that are beyond the scope of this thesis' inquiry. The extent to which foreign investments can affect positively the performance of firms and ultimately that of entire host economies is as far as the estimated results in this thesis tend to suggest, ceteris paribus.

The innovation in this thesis is twofold. Firstly, this thesis employs a wide range of methodological techniques such as cluster analysis using machine learning that have hardly been used before by earlier studies in this type of analysis. I also employ conventional regression but also other traditional impact evaluation approaches in my analysis as robustness checks for validation of some of the results, particularly in Chapter 4. Secondly, I utilise three unique but comparable comprehensive micro data sets in the analyses and extend the empirical investigation to examine the likely effect of foreign ownership on firm performance in SSA, this being a region of limited focus in earlier studies.

Chapter 2

Foreign Ownership and Private Enterprises in Developing Economies

A Cluster Analysis of Ugandan Firms

2.1 Introduction

Foreign-owned and domestically-owned firms, in a typical host economy environment, are on average, different. Foreign-owned firms for instance, have been associated with higher export levels, higher average wages, intensive expenditure on research and development, high employment levels, larger firm sizes; and they are generally more productive than domestically-owned firms (Kimura & Kiyota, 2006) (Erdal & Göçer, 2015a). These generally positive firm characteristics make foreign-owned firms particularly attractive to policy makers. As indicated in chapter 1, much of the existing research on foreign-owned investments at the firm level does not adequately answer the questions of; whether foreign ownership causes these differences, whether foreigners acquire better-performing firms or whether comparing foreign-owned and domestically-owned firms in fact compares different types of firms in which foreign ownership cannot be associated with the observed differences in outcomes. This chapter investigates this last point in more detail: whether the observed differences between foreign-owned and domestically-owned firms are due to comparisons between different sub-groups, such as between large firms (where foreign ownership is more prevalent) and small firms, or whether foreign-owned firms are systematically different from domestically-owned firms across a number of dimensions.

In this investigation, this chapter contributes to the growing literature on foreign-owned investments by using a relatively novel method of analysis, in a developing country – Uganda – where analysis of foreign-owned investments has been limited. Suitability of Uganda as a case in this thesis was partly derived from the fact that, besides data availability, her key domestic economic policy frameworks have embraced the promotion of private investment with a focus on incentivising foreign-owned investment over domestically-owned investment. However, just like the SSA situation alluded to in chapter 1, the anticipated economic outcomes in Uganda are still dismal. Economic outcomes such as low export volumes averaging only

$121.45 million over the period 1993 to 2017[1], youth unemployment documented to be at 19.7% by 2011/12 [2] or even higher at between 64% and 70% by 2014[3] are still prevalent. Over 80% of the manufacturing industrial sector is still made up of firms of small size that produce goods of low value. The manufacturing sector contribution of only 7% of gross domestic product (GDP), which is below the average of 11% for less-developed countries (LDCs), and also below the East African regional average of 10%[4]also attest to the observation above. These outcomes make Uganda a suitable case that is reflective of the SSA economic outlook pointed out in chapter 1. Amidst increasing foreign-owned investment inflows into Uganda over the years, the outcomes abovementioned reflect less of the characteristics associated with foreign-owned firms as also highlighted in chapter 1. They equally imply that foreign investments have had less of a contribution to the country's welfare as opposed to what is expected and commonly noted in the empirical literature about such investments. This anomaly makes scholarly inquiry into firm-level features of private investments in Uganda valid, timely, and hence appropriate in the SSA context as a case study. Evidence from analysing Uganda's case will reveal whether the kind of private foreign investments attracted and incentivised mirror the conclusions in literature; and with the dismal economic outcomes thus potentially being attributable to probably other factors. Moreover, findings will set a firm stage for a much broader inquiry about these investments in SSA. In Uganda just like probably in comparable economies in SSA, less evidence is available on the superiority or abilities of foreign-owned firms relative to domestically-owned firms since relatively little firm-level research has been conducted there, especially regarding firm characteristics in relation to or associated with ownership. Moreover, the scanty evidence sometimes shows conflicting findings. Moss and Ramachandran (2004), and Wilson and Cacho (2007), for example, find that foreign-owned firms are, on average, larger and more productive than domestically-owned firms. However, Niringiye and Tuyiragize (2010), and Obwona (1998) find most firms to be small and medium. This scarcity of evidence limits not only frontiers of knowledge but also fails to provide much-needed guidance for policy development.

[1] Trading Economics 2017 Uganda chapter
[2] Uganda Bureau of Statistics 2015 Statistical Abstract- page 28
[3] Magelah and Ntambirweki-ACODE (2014)- Youth unemployment and job creation in Uganda: Challenges and Opportunities, page ii
[4] African Development Bank Report of Eastern Africa Manufacturing sector, promoting technology, innovation, productivity and linkages,2014, pages 2-4

In terms of approach, unlike previous studies that are based on *a priori* assumptions, in this chapter I use cluster analysis, which reverses the process by allowing the data to classify firms. The characteristics are then analysed based on group classifications. This methodological approach allows this chapter to have an additional contribution as a robustness check on earlier studies and their conclusions. Cluster analysis is a strong tool of multivariate exploratory data analysis. It is effective in aggregating objects like firms based on their characteristic similarities (Murry, 2016). It involves a great number of techniques, methods, and algorithms that are applicable in various fields, including economics (Řezanková, 2014). Its aim is to identify groups of similar objects (households, firms or countries) according to selected variables like employment and poverty indicators (Řezanková, 2014). Through data exploration, cluster analysis is useful in generating hypotheses and in testing their presence in the data; and also, in testing groupings developed by other studies using other methodologies. It is also useful in the development of typologies and investigation of useful conceptual schemes for grouping entities (Gries, 2007).

Historically, cluster analysis was firstly used within the disciplines of biology and ecology (Yim & Ramdeen, 2015: 9). Although this technique has been utilised in the field of social sciences, it has not gained as much widespread popularity as in the natural sciences. Some of the fields in which custering techniques have been applied are; marketing (Mooi & Marko, 2011), clinical medicine and health sciences (Fersini *et al.*, 2012), and machine learning, pattern recognition, image analysis, information retrieval and bioinformatics (Bijura, 2013). In the available literature, cluster analysis is also referred to as mathematical taxonomy or unsupervised classification. It has also been referred to as segmentation and partition analysis, although these terms are often contested[5] (Tryfos, 2002). A general interest in cluster analysis increased in the 1960s, leading to the development of several new algorithms that expanded possibilities of analysis. Moreover, in recent decades, there has been gradually increasing incorporation of cluster analysis in the areas of health and social sciences (Yim & Ramdeen, 2015: 9). However, despite this growth, the application of cluster analysis in economics gained traction only recently (Michael, (2016).

Therefore, the work in this thesis and this chapter in particular extends the application of cluster analysis in economics and makes use of the World Bank Enterprise Survey data to classify

[5] Irrespective of the name, all the techniques aim at identifying groups in the data by classifying cases into groups that are relatively homogeneous within themselves and relatively heterogeneous between each other.

Ugandan firms based on performance indicator variables. The major aim is to establish whether the observed differences between foreign-owned and domestically-owned firms are due to comparisons between different sub-groups (for example, between large firms where foreign-ownership is more prevalent and small firms where foreign-ownership is less prevalent); or whether foreign-owned firms differ from domestically-owned firms systematically and significantly across a number of dimensions. This will help to confirm or contradict earlier studies on the characteristic features of private firm enterprises in Uganda and also set the stage for broader analysis of the same on SSA. These study findings will further provide a blueprint for economic policy reorientation not only in Uganda but also in comparable developing economies in SSA.

This study is the first to analyse firm-level data for Ugandan firms using cluster analysis and application of ClustOfVar. The latter is a relatively novel algorithm for dimensionality reduction on various types of data, including mixed data. This approach has, however, been previously employed successfully elsewhere by other scholars, such as Naughtin and Rankin (2016), on South African manufacturing and exporting firms.

The analysis of characteristics of firm clusters in this chapter shows a group or cluster of firms that is more likely to differ from other clusters on a number of performance indicators. Relative to other groups, firms in this cluster are generally likely to: be larger in size, exhibit higher labour productivity, to spend on R&D, be capital intensive, possess relatively experienced management, offer formal training to workers, and export part of their output. Above all, firms in this cluster are, on average, associated with higher foreign ownership stake, pointing to the likelihood that such a cluster is composed of mainly foreign-owned firms. When variables are classified also, and their cluster characteristics analysed, results indicate the following: greater association between ownership status, employment, and experience within the same cluster; moderate association between ownership and variables such as exports, worker_educ, and capital_intensity; and almost zero correlation with wage, labour productivity, and material use. When a classification and regression tree analysis was performed, ownership status, worker_educ, management experience and labour productivity were revealed as some of the important variables that sort firms into particular clusters.

The remainder of this chapter is laid out as follows:

- Section 2.2 presents the theoretical background on foreign direct investments;

- Section 2.3 presents the empirical literature review on characteristic determinants of foreign direct investments;
- Section 2.4 contains a discussion on the theoretical underpinnings of cluster analysis;
- Section 2.5 presents an empirical literature review on cluster analysis;
- Section 2.6 presents a description of the methodology, data and the results; and
- Section 2.7 contains the study conclusions.

2.2 Theoretical background

There are a number of theoretical explanations of foreign direct investment. Dunning (2003: 176), Feng (2017: 1), Faeth (2009), Hosseini (2005) summarise this well. Some of these theoretical explanations are discussed in this section and form the theoretical guide for analyses in this chapter. In the early attempts of theoretically explaining foreign direct investment, the Heckscher-Ohlin (HO) neoclassical theory was the most dominant (Faeth, 2009: 166). The HO model was based on the assumption of a fixed endowment of capital and labour for each country, with each country producing a mix of products that made the best possible use of its resource endowment, access to homogeneous technologies by each country and the existence of perfect markets for both inputs and goods. It further assumed only two countries, constant returns to scale production functions and zero transport costs. In this theory, trade is driven by the differences in endowments across countries and relative differences in commodity factor-use intensities. Therefore, a country exports goods which, in their production, use the country's abundant factor intensively. For example, countries with surplus capital would export capital-intensive goods. In the absence of commodity trade, firms would move capital to foreign capital-deficit economies where higher returns are anticipated. Movement of capital in this case is what gives rise to a foreign-owned firm (popularly referred to as foreign direct investment firm in literature) in another country.

Associated with the HO theory were theories such as MacDougall(1960) and Aliber (1970). MacDougall assumed perfect competition, constant returns to scale, full employment of resources, two factor inputs and only one good. Both MacDougall (1960) and Aliber (1970) agree that capital moves to where returns are higher, as with HO. Aliber's departure is in the cause of differences in returns to capital, which, according to him, emanates from not only capital endowments but also currency risks. Such risks are associated with the interest premium charged inclusive of the expected currency depreciation. Foreign-owned firms, which are usually from less volatile currency areas, can borrow from countries with more volatile currencies to finance similar investment streams more easily than domestically-owned firms in host economies. This advantage alone acts as a stimulant for foreign-owned firms to invest in host economies, hence growth of foreign direct investment. Given the HO and associated theories' assumptions, flows and growth of foreign direct investment were more likely determined by differences in resource endowments between origin and host countries; differences in relative factor intensities of goods produced; and differences in currency risks between economies.

The HO theory, however, is featured by some limitations in explaining the emergence and operations of the foreign-owned firm in relation to the domestically-owned firms in a given economy. One such limitation is the assumption of perfect competition, which is unrealistic. Moreover, absence of transaction costs in this theory leaves no ground to distinguish between direct and indirect investment (Hosseini, 2005: 531). Also, economies can manipulate returns on capital and flows of capital by imposing taxes on international capital movements to improve their welfare. Using this theory, it is further difficult to account for foreign-owned investment flows between countries with unified currency areas, such as the European Union (EU). The above limitations motivated Kindleberger (1969) and Hymer (1976), two of the first scholars to criticise the neoclassical approach and HO theories, to advance the theory of ownership advantages as an alternative.

The main argument and/or fundamental assumption of the theory of ownership advantages is that foreign direct investments can only flourish under structural market imperfections. Based on the 'monopolistic advantage' hypothesis, Hymer and Kindleberger posit that distortions are necessary in international markets if foreign direct investment is to arise. To this end, foreign-owned firms must have specific ownership advantages that give them a competitive edge over domestically-owned firms in countries where foreign-owned firms intend to start production operations. These advantages, which must outweigh the penalties of being foreign include product differentiation (imperfect good markets), managerial expertise, new technology or patents (imperfect input markets) (Denisia, 1998), the existence of external or internal economies of scale (Faeth, 2009: 167), superior access to capital (Kindleberger, 1969) and superior brands (Caves, 1971: 6).

Two notable strands support this hypothesis with minor variations. Firstly, Caves (1971) focuses on the advantage of product differentiation to substantiate the likelihood of horizontal foreign-owned investment by arguing that foreign direct investment is the best option, other than licensing or exportation, once knowledge is used to differentiate products other than managerial skills. The second strand is the oligopolistic reaction perspective put forward by Knickerbocker (1973) who proposes that, although firms operate in imperfect markets and foreign direct investment flourishes under such conditions, the basis is an oligopolistic reaction, especially the 'follow the leader' strategy or reaction to foreign firms 'invading' the foreign-owned firms' home market (Faeth, 2009: 168). Based on the above assumptions and from the perspectives of the monopolistic and oligopolistic hypotheses, the theoretical prediction is that, once firms have these advantages outweighing those of domestically-owned firms, they will be motivated to invest directly in the foreign country. This motivation is driven by the desire to reduce international competition among firms and the need to increase gains from firm-specific characteristics or advantages (Hosseini, 2005). Visible in this theory, too, is the centrality of firm-specific characteristics and the imperfections in international markets in any attempt to analyse flows and growth in foreign direct investment. This centrality of firm-level characteristics or advantages points to the likelihood of systematic differences between foreign-owned and domestically-owned firms even prior to foreign investors setting foot in the host economy. An investigation of whether this theoretical prediction holds for Uganda and indeed SSA is what empirical work in this thesis attempts to confirm.

Despite its apparently strong explanatory power of the foreign direct investment firm, this theory doesn't justify why firms opt for direct investment rather than licensing their ownership advantages to domestically-owned firms abroad and earning economic rents through licensing fees while sidestepping the costs of doing business abroad. In other words, this theory falls short at explaining the desire for direct involvement or a visible hand in business abroad. Elsewhere, this theory falls short at explaining why only some firms have these ownership advantages and others don't; and how those that actually have, obtained such advantages. As put forward by Buckley and Casson (2009), the answer to these interesting questions, lies partly in the fact that the cost of licensing may even be higher than the costs associated with business operations abroad. The above limitations of Hymer and Kindleberg's theory elicited further scholarly response to overcome them, which response gave birth to one of the notable post-Hymer theories, the internalisation theory.

The internalisation theory is associated with the works of Buckley and Casson (1976), Casson (1983) and Rugman (1981). According to Feng (2017: 3), this theory integrates Coase's (1937) theory of the firm with theories of international trade and economic geography, such as those described by Weber (1929) and Ohlin (1933), to formulate a theory of multinational corporations. This theory is also cognisant of the fact that firms operate under imperfect markets. Its fundamental assumption is that of rational action: rational agents will internalise markets when the anticipated returns are more than the associated costs; and profit-oriented managers of a firm will internalise intermediate product markets up to the point where the costs and benefits of internalisation cancel each other out (Buckley & Casson, 2009: 5). Other key principles of this theory are that a firm's profitability and the dynamics of its growth are based upon a continuous process of innovation resulting from Research and Development; and that firms seek out the lowest-cost location for each activity, taking the activity's linkages with other activities into account (Buckley & Casson, 2009: 6). This principle highlights, R&D, one of the likely key distinguishing firm-level features between foreign-owned and domestically-owned firms and which has implications for firm-level performance.

Hymer and Kindleberg's theory indicated that the ownership advantages of firms are subject to imperfect markets; and that these advantages include production and marketing technology, marketing skills, R&D, and services. They are, however, marked by high risk and uncertainty, resulting into high transaction costs in terms of; enforcement, information, and bargaining costs according to Faeth (2009: 168). They are further subject to difficulties of appropriability due to their 'public good' nature, for example, technology (Magee, 1977 as quoted by Faeth (2009)). These features make the transaction costs associated with the ownership advantages very high. If left to the whim of the market mechanism, such costs keep rising. By rational action, the rise in transaction costs motivate profit-maximising firms to eliminate such imperfections and costs through either the internalisation of the flow of knowledge originating from R&D undertakings or by the internalisation of operations involving the flow of intermediate products through sequential production stages and channels of distribution. With internalisation, the activities that were previously linked through the market mechanism (an invisible hand) are brought under control and common ownership in a "market" internal to the firm (a visible hand) (Buckley & Casson, 2009: 2).

The internalisation process can lead to direct investment in a foreign economy because it can occur either within or across national boundaries. Once internalisation takes place across national boundaries, a foreign direct investment firm is born, arising from the firm's desire to get involved directly in business management as opposed to via a licensing model. For instance, knowledge internalisation stimulates foreign direct investment because of what is known as 'buyer uncertainty'. A knowledge-generating firm would easily license such knowledge generated through R&D to a producing firm anywhere in the world. However, licensing can be a nightmare because of potential payment default arising from probable suspicions about lack of novelty of generated knowledge, the ability of the buyer in a foreign country to invent around the novel technology, and the likelihood of the buyer selling the knowledge to the competitors of the knowledge generating firm. These factors can lead to direct investment being the optimal choice. From this perspective, the theoretical prediction of internalisation easily points to more foreign direct investment presence in knowledge-intensive industries and sectors. Another predictive strand of this theory is from Hennart (1982, 1991), whose theorisation is derived from McManus's (1972) theory of property rights and Williamson's (1975) 'market and hierarchies' (Faeth, 2009: 169). According to Hennart, internalisation advantages can emanate from a firm's reputation or know-how – leading to horizontal integration – or failure in other markets – leading to vertical integration. The above cases of firm reputation or market failure would stimulate foreign direct investment in a bid to replace the price system of the market in order to gain efficiency and avoid transaction costs.

The internalisation theory unlike earlier theories reveals the theory of the foreign direct investment firm as a distinctive case of a general theory of the firm that embraces both multinational and domestic firms (Buckley & Casson, 2009: 11). It explains why firms are not willing to license their ownership advantages, such as technologies to others in foreign economies, and gain rents. It further explains why foreign direct investment is mainly found in knowledge-intensive industries; and supports the empirical postulations that foreign-owned firms are characterised by higher levels of R&D expenditure, high advertising expenditure and the employment of a highly skilled labour force. It further supports earlier theories that firm ownership advantages and characteristics are key determinants of foreign direct investments. In this theory, entrepreneurs pinpoint profit prospects and then build viable global markets, supported by global production systems and a commitment to ongoing research and development (Buckley & Casson, 2009: 11).

Despite its explanatory power, the internalisation theory doesn't explain the timing of foreign direct investment, its subsequent expansion or the choice of country of location, amongst other shortcomings. For instance, after World War 2, European countries were, unlike economies in Africa, popular destinations of foreign direct investment, yet for sure domestically-owned firms in Africa were more inferior (compared to those in Europe) to foreign direct investment firms and little in number, which would ideally mean more foreign direct investment inflows to Africa. Currently, the Asian and BRICS (Brazil, Russia, India, China and South Africa) countries attract more foreign direct investment than African economies, with the exception of South Africa.

The limitations of the internalisation theory and previous theories to fully account for variations in foreign direct investment flows around the world were addressed by Dunning (1977, 1979) in a new theoretical formulation that focused on the location specifics of foreign direct investment firms, while at the same time incorporating arguments on both firm-specific advantages and internalisation. Dunning's theory, known as the 'Eclectic paradigm', is based on four principles or assumptions that must be fulfilled by any firm intending to internationalise (Dunning, 1993),(Dunning, 1993a: pp 79) (Stoian & Filippaios, 2008). These principles are;

1) **ownership advantage**: the firm must have specific ownership advantages compared to firms located in the country where it intends to establish new plants;
2) **internalisation advantage**: the benefits of internalisation must offset the costs or those benefits associated with other alternatives of serving the foreign market, for instance exportation and licensing;
3) **location advantage**: it must be beneficial to utilise the firm's ownership advantages abroad rather than at home in concurrence with domestic resources in the foreign country; and
4) the extent to which the firm feels internalization is consistent with its long-term management strategy

The first three principles listed above are also referred to as the "OLI" (ownership-location-internationalization) framework.

Ownership advantages in the OLI framework are enshrined in the firm's production process, ensuring a competitive advantage over domestically-owned firms. These include patents, management know-how, technical knowledge and reputation. The more competitive advantages, the more the firm is motivated to invest directly in the foreign economy (Dunning, 1980). Location advantages that are motives for producing in the foreign country include lower costs of production and transport, favourable tax regimes, access to protected markets and low domestic competition. As mentioned earlier, internalisation arises from the 'public good' attributes of ownership advantages and is motivated by the need to lower transaction costs, minimise technology imitation and control the firm's brands and reputation through effective management and quality control (Faeth, 2009: 171). This theory further distinguishes the ownership advantage element into two sets: those related to proprietary assets owned by the firm whose internalisation is a choice of the firm; and those ownership advantages whose exploitation is only possible if internalised. According to Dunning, the latter are transaction-ownership advantages, while the former are asset-ownership advantages.

The triumvirate underpinning the eclectic paradigm (of ownership, location and internalisation) highlights the theoretical predictions that are associated with this framework regarding the determinants of flows and growth of foreign direct investments around the world, from origin to host economies. These predictions are dependent on whether the enquiry is focused on location or internalisation; ownership; countries, firms or industries; or even different forms of foreign direct investment. It thus provides a comprehensive framework that suits any form of inquiry including the key focus of this study, being firm-level characteristics.

Among other predictions, the L component in OLI is likely to be influenced by market size, labour costs, human capital, economic openness, cultural affinity, the frequency and likelihood of government change, natural resources, infrastructure, population, property rights protection and fiscal policies. The ownership and internalisation aspects of the firm that intends to internationalise are as discussed in the associated theories. In the OLI framework, the ownership advantage involves the following: technological strengths; monopolistic power, economies of scale; and internalisation advantage that organises productive activities within a firm across national boundaries; and the location advantage (this being an area where determinants of foreign direct investment are abundant) (Feng, 2017: 5). However, the OLI advantages vary depending on whether the country is developed or developing, small or large, industrialized or not; additional factors are whether firms are large or small, innovating or imitating, technologically limited or advanced, competitive or monopolistic, and assembling or processing (Dunning, 1988).

Although most scholars concur that the eclectic theory is the most comprehensively applicable theory for foreign direct investment analysis, it has also been criticised for a number of reasons. It has many variables, making analysis complicated; it exhibits bias against developing economies in which firms lack the ownership advantages; and it falls short in not spelling out the menial role of financial aspects in the foreign direct investment decision (Forssbeck & Oxelheim, 2008). But its efficacy in aiding foreign direct investment analysis at both micro and macro level is thorough; and it provides theoretical support for the fact that imperfections in international markets dictate flows of foreign-owned firm activities, which in turn underline foreign-owned firms' likely *ex ante* differences from domestically-owned firms on several dimensions. Dunning's theory and those discussed thus far do not account for the entire spectrum of a firm's desires to invest abroad. Further elaboration of this spectrum can be found in the factor-proportions and proximity-concentration theories.

The factor-proportions theory attempts to explain movement and growth of foreign direct investment from the perspective of a firm's desire to integrate production vertically across national boundaries in order to take advantage of factor-price differences associated with different relative factor supplies (Brainard, 1993a). Under this theoretical framework, trade costs are set to zero, firms are assumed to prefer producing in one location due to increasing returns to scale, internalisation of production is taken as given, and information regarding factor endowments is asymmetric. Based on these assumptions, the factor-proportions hypothesis predicts that firms will establish production units abroad once differences in terms of factor endowments between home and foreign economies are large and when factor price differences actually exist. Specifically, under asymmetric information on factor endowments, firms from human-capital-surplus economies will be stimulated to establish units in human-capital-deficit economies, giving birth to vertical foreign direct investments. Exporting services (like R&D and advertising) and intermediate goods from foreign direct investment headquarters to new production units in foreign markets makes intra-firm trade a key feature of vertical integration Firms are stimulated to establish the production unit in a foreign economy in order to avoid duplication of their knowledge capital, this being a basic advantage they have over domestic firms. With knowledge capital and its joint-input attribute, the operation of multi-plants ensures growth in firm-level scale economies (Faeth, 2009).

The proximity-concentration theory seeks to explain flows and growth of foreign direct investment from the perspective of a trade-off between maximising nearness to customers and the concentration of production in one geographical place so as to enjoy economies of large-scale production. This theoretical thinking accounts for the emergence of horizontal foreign direct investment across boundaries of economies. It is associated with the scholarly works of Krugman (1983), Horstmann and Markusen (1996), Brainard (1993a), and Markusen and Venables (2000). This theory assumes that there are firm-specific costs, tariffs, and transport costs that motivate firms to engage in both domestic and foreign production. It also assumes that there are plant-scale economies that incentivise firms to produce at home only and to export to the foreign market (Faeth, 2009: 175). In terms of theoretical prediction, the proximity-concentration hypothesis posits that firms are more likely to establish production units in foreign economies horizontally if transport costs are higher, trade barriers are strictly in place, investment barriers are low, and the size of economies of scale at firm level is low relative to plant level (Brainard, 1993b:520). This also implies that foreign-owned investments are more likely to exist in industries and sectors with large firm-specific costs along with high tariff and transport costs; and that they are less likely to be found in industries or sectors with relatively small plant-scale economies. Logically, a firm wouldn't set up a plant abroad when transport costs are negligible (meaning exportation makes sense economically) yet it has to incur fixed costs in setting up such a plant in the first place, but without saving significant costs in terms of transport. Hence, high transport costs, trade barriers and large-firm-level economies *vis-à-vis* plant economies are what provide the logical motivation for firms to go foreign. Intuitively, the proximity-concentration hypothesis predicts the emergence of horizontal foreign-owned firms as a consequence of possible substitutability between exportation and foreign production.

The factor-proportions and proximity-concentration theories are not without their inherent limitations in explaining emergence and growth of foreign direct investment. The proximity-concentration hypothesis, for instance, predicts the growth of horizontal foreign direct investment on the premise that different activities use either factor inputs in the same proportion or only one factor input. In such a situation, there is no factor price justification for vertical foreign direct investment. A key assumption of the factor proportions hypothesis is the absence of trade costs, which eliminates any reason for horizontal foreign direct investment (Anghel, 2007: 5). Horizontal foreign direct investment is also predicted as likely to occur between countries at the same stage of development, which is far from reality. These and other shortcomings played a vital role in motivating Markusen and other scholars to come up with the Knowledge-Capital Model of foreign direct investment.

The KCM is an integration of both the factor-proportions and proximity-concentration models of foreign direct investment in a single general equilibrium model. It allows for the building of multiple plants and the separation of headquarters services, like R&D, advertising and production as distinct (Faeth, 2009). Unlike the factor-proportions theory, the KCM assumes the existence of trade costs between economies, hence allowing for both low and high costs of transport between countries. Unlike the proximity-concentration hypothesis, the KCM assumes dissimilar factor-inputs intensities across activities (Anghel, 2007: 5). Other assumptions derived from its early formulation include a 2x2x2 formulation with one good exhibiting constant returns to scale and the second with plant and firm-level economies of scale and variations in country size.

The KCM is a result of the works of Markusen (1997; 2000), in which horizontal stimuli of foreign direct investment are combined with vertical stimuli. Horizontal stimuli are associated with firms' desire to take production nearer to customers and elude trade costs, while vertical stimuli express the desire to engage in unskilled, labour-intensive production activities in areas with relative surpluses of unskilled labour. Based on this view, similarities in market size, resource endowments and transport costs are some of the causes of horizontal foreign direct investment, whereas vertical foreign direct investment is much better explained by differences in relative resource endowments (Faeth, 2009: 179).

Based on these assumptions, KCM's theoretical prediction is far more accommodating than other hypotheses, with firms having alternatives of either geographically separating their headquarters from a single plant or setting up numerous plants. In both cases, there are possibilities of firms integrating horizontally or vertically, or even deciding to remain domestic and to serve foreign markets via exportation. It turns out vividly that the underlying factor for either vertical or horizontal foreign direct investment emergence is possession of knowledge capital, which, if not possessed by a firm, relegates it to the domestic market, with this pointing to how foreign-owned firms differ from domestically oriented firms even in the countries of origin.

As indicated earlier, theoretical explanations of foreign direct investment are numerous. While earlier attempts by the neoclassical theory received criticism due to its assumptions related to perfect competition, several other theories have been developed based on imperfect market analyses. Theories like those based on internalisation, ownership advantages, product life cycle formulations, vertical or horizontal integration formulations, and policy variable theories appear in the vast foreign direct investment literature, all with the single goal of explaining the motivating factors behind the flows and growth of foreign direct investments. However, by scholarly consensus, Dunning's OLI theoretical formulation remains a strong approach for explaining foreign direct investment. His theory combines ownership, location, and internalisation advantages as key determinants of foreign direct investment. Ownership and internalisation were highlighted by previous theories and more theoretical modifications continue to emerge in modern foreign direct investment analyses. As put forward by Faeth (2009: 188), analysis of foreign direct investment should not be based on a single theory but rather more broadly on a combination of ownership advantages, such as market size, agglomeration economics, market characteristics, cost factors, risk and policy variables. Numerous empirical studies have adopted this approach, even when analysing specific aspects or theories of foreign direct investment. This study will also follow this approach in its attempt to analyse systematic differences between foreign-owned and domestically-owned firms in Uganda and SSA generally. This theoretical analysis provides firm guidance in selection of key variables to use in the analyses.

2.3 Empirical Literature Review

The theoretical review in the previous section has been enhanced by empirical research findings. All the theories discussed in the previous section have been empirically tested in various scholarly works. Table 2.1 provides a summary of some of the empirical studies in which these theories have been tested and evaluated in terms of their ability to explain the features and flows of foreign direct investments around the world. In terms of scope, relatively more studies have been at micro level. At the micro level, studies have exploited firm-level data to investigate the extent to which theoretical formulations hold in as far as accounting for foreign direct investment growth is concerned. Macro- and industrial-level approaches have used industry-level data and macro variables to explain the motivations of firms engaging in international production. In terms of geographic scope, most studies have been carried out in affluent economies, with only a few similar studies having been undertaken in developing economies. In terms of methodology, most empirical studies have employed conventional econometric methods, mainly using regression techniques, with very few applying alternative techniques.

Table 2. 1: Selected empirical studies on characteristics and determinants of FDI

Author (s)	Country	Method	Data
Correa da Silveira *et al.* (2017)	Brazil	Logistic Regression	Firm-level
Ablov (2015)	Poland	Pooled OLS	Firm-level
Leman and Ismet (2015)	Asia	Panel regressions	Firm-level
Wakasugi *et al.* (2014)	Japan	Descriptive statistics	Firm-level
Feng-Jyh Lin (2010)	China	Hazard modelling	Firm-level
Falk and Wolfmayr(2010)	Austria	Matching and DID	Firm-level
Cai and Guney (2010)	China	FE and GMM (Panel)	Firm-level
Lee, Huang and Chan (2009)	China	Logistic regressions	Firm-level
Barther *et al.* (2008)	Ghana	OLS	Firm-level
Mayer and Ottaviano (2007)	Europe	Gravity regressions	Firm-level
Raff and Ryan (2006)	Japan	Hazard models	Firm-level
Kimura and Kiyota (2006)	Japan	Regression	Firm-level
Moss and Vijaya (2005)	East Africa	Regression	Firm-level
Love and Lage-Hildago (2000)	Mexico/US	Cointegration analysis	Time series
Rubio and Rivero (1994)	Spain	Cointegration analysis	Time series

Source: Author's own compilation based on literature review

The standard neoclassical theory's ability to explain flows and distribution of foreign owned investments across economies has been empirically tested. Harm Zebregs (1998) investigates whether the observed distribution of foreign direct investments across developing economies can adequately be explained by the neoclassical theory or by its extended modifications in which technologies are allowed to vary across countries. Zebregs estimates production functions that fulfil the assumptions of the neoclassical theory as constant returns to scale and diminishing marginal returns to capital and labour. Marginal rates of return to capital are computed on the basis of the estimated production functions and compared with the pattern flows of foreign direct investments across developing economies. The findings suggest that the marginal product of capital as defined by the neoclassical theory cannot explain the distribution of foreign direct investments in developing countries. If the neoclassical theory is to be taken in its form, with only labour and capital, and with technological homogeneity across economies, then poorer economies should have large foreign direct investment inflows due to their high rates of return on capital. However, this is not the case.

The ownership advantages theory has been empirically tested, with results revealing foreign direct investment firms being characteristically unique and possessing ownership advantages when compared with domestically-owned firms, irrespective of whether the inquiry is focused on firms from the same country (where foreign firms are outward – as opposed to inward – investment) or not. Regression and other associated methods have been popularly used, and ownership advantages have been confirmed to be characteristic determinants of foreign direct investment. The earlier studies first provided evidence to Hymer and Kindelberg's monopolistic hypothesis. One such study by Horst (1972), based on data from over 1000 manufacturing firms in the US, analysed the market share of United States (US) industries in Canada. A key result from Horst's analysis was that most foreign direct investments are of larger size than comparable domestically-owned firms. This result resonates with the empirical analyses conducted at industry level, especially in the developed world. For instance, Wolf (1977) analysed industry-level data on US manufacturing firms using regression techniques. Wolf's findings indicated that sales of US foreign affiliates as a percentage of domestic production were increasing in average firm size and technical manpower (Faeth, 2009: 169). Further industry-level evidence on other features besides firm size is visible in Lall's (1980) study using data from 25 US industries involved in foreign production and exportation. Key results of this study were the significant increases, for foreign affiliates of US firms, in expenditure relating to R&D, advertising, and average wage for employment. These results concur with those of a later study by Saunders (1982), whose analysis found advertising, R&D, and managerial resources to be significant characteristics of foreign direct investments. Other empirical studies, using cross-sectional data from primarily the US and Sweden acknowledged the above theoretical underpinnings, although to varying degrees. For instance, Blomstrom and Lipsey (1986) found that firm size had only a threshold effect on foreign direct investment but with either less or no effect thereafter. Studies like Blomstrom and Lipsey (1986) highlight one other element noticeable in foreign direct investment literature, namely, the variance in empirical results regarding significance associated with ownership advantages of foreign direct investment firms.

Elsewhere, focusing on a few ownership advantages, Falk and Wolfmayr (2010) apply matching methods and difference-in-differences estimation technique to investigate the features of Austrian firms. Their study found Austrian firms that internationalised as outward foreign direct investments to be significantly different from Austrian firms that were domestically-oriented. Firms that engaged in foreign direct investment were found to be mostly large in terms of employment and sales figures, capital intensive, they paid higher wages and were more likely to hire highly educated workers. These findings, which are also found to hold for inward foreign direct investment firms from the rest of the world, point to the conclusion that "firms that start foreign activities are *ex-ante* different from non-investing purely domestic firms" (Falk & Wolfmayr, 2010: 1). Before firms decide to directly invest abroad instead of licensing or exporting, several firm-specific characteristics determine such a decision.

In a methodologically different study, Ablov (2015) applied OLS and panel-data econometric methods on the firm-level data of 147,878 Polish companies from all sectors of economy to investigate firm-level characteristic determinants of foreign direct investment. The results indicated that firm productivity, R&D, firm size, and highly skilled workers were significant features of foreign direct investment firms that established investments in Poland (Ablov, 2015: 91). Lin (2010) adduces further evidence in her investigation of Taiwanese foreign direct investments in China's IT sector from 1996 to 2005. Using firm-level financial data and employing event history techniques under Cox's maximum likelihood proportional hazard modelling, Lin finds export orientation, capital intensity, firm performance, R&D, and firm size positively influencing firms' engagement in foreign investment. By implication, before internationalisation, firms are characterised by ownership advantages that not only stimulate them to invest in foreign markets but also distinguish them from domestically-owned enterprises located in economies where such foreign-owned firms intend to invest.

Focusing on capital intensity, the ownership advantage theory has further been empirically tested in the recent past. Among recent scholars with such specific evidence are; Cipollina, Pozzolo, Giovannetti & Filomena (2012), Cole and Elliot (2005), and Aaron, Iregui & Ramírez (2014). Visibility of foreign ownership in capital-intensive sectors is one argument that has been used to support the conclusion that foreign-owned firms are relatively more capital intensive than domestically-owned firms and it is a source of competitive advantage. One other explanation that has been put forward in support of foreign-owned firms being capital intensive is the fact that the decision to invest directly in a foreign economy is traded off with that of continued exportation to the foreign economy. The former alternative involves greater sunk costs, implying that the firm must have reached a given level of productivity and volume of capital (Melitz, 2002; Raff, Michael & Staehler, 2006). This tends to imply that foreign direct investment firms are capital intensive *ex-ante*. Lin (2010) and Siddharthan and Nollen (2007) conclude that it is capital intensity that gives foreign-owned firms superiority in terms of product quality and cost advantages in exportation. Aaron *et al.* (2014: 8) apply panel probit modelling to study the determinants of foreign direct investment in Columbia. Their findings suggested that firms that listed on the stock market, engaged in foreign trade activities, and operating in sectors with higher capital intensity were more likely to be recipients of foreign investment. Specifically, study findings revealed a 0.3% marginal effect on foreign participation arising from increased capital intensity in largely social and community sectors of Columbia. This affirms capital intensity to be an important characteristic of firms with foreign ownership. These study findings resonate with those of Karpaty and Poldahl (2006) in Swedish firms and scholarly findings by Giulietti, McCorriston and Osborne (2004). The capital intensity of firms is pertinent in determining foreign direct investment levels as the size of the resource commitments required to participate in foreign investment can vary substantially depending on how capital-intensive firms are (Lin, 2010). Several other studies that have recently investigated the interlinkages between firm and industry-specific characteristics and foreign direct investment include; Hagemejer and Tyrowicz (2012), Wang, Alba and Park (2013), Liu and Nunnenkamp (2011), Raff and Ryan (2008), Hilber and Voicu (2010). These have provided further evidence of the significance of firm ownership advantages and foreign direct investment, highlighting the dichotomy between foreign-owned and domestically-owned firms.

The ownership advantages theory has also received some attention in developing economies more recently. Moss and Ramachandran (2004) uses the World Bank's Regional Program on Enterprise Development (RPED) data on 300 - 400 manufacturing firms in Kenya, Uganda and Tanzania, providing further evidence of ownership advantages for foreign-owned firms. Using regression techniques, their analysis found foreign-owned firms employing more workers, likely to offer formal training programs to their workers and registered higher value added per-worker than domestically-owned firms. In Uganda, for example, the average number of workers for foreign-owned firms was found to be nine times that of domestically-owned firms, with an average of 20% more foreign-owned firms likely to offer formal training to their workers when compared to domestically-owned firms.

Barthel, Busse and Osei (2008) employed multivariate analysis on the World Bank Enterprise Survey data for Ghana in an attempt to investigate the salient characteristics and determinants of foreign direct investments. Their investigation also investigated how domestically-owned firms differ from foreign-owned firms. Key amongst their study findings was that foreign-owned firms in Ghana are not likely to have more experienced management and are not likely to invest or export more than their domestically-owned counterparts (Barthel *et al.*, 2008: 9). These findings deviate from earlier studies that have investigated ownership advantages of export intensity and management ability of foreign-owned firms. Whether this contradiction in empirical study findings has to do with the level of development of the economy in question or not is not conclusively answered by my literature review. However, in terms of firm size, Barthel *et al.* (2008) are in agreement with existing studies in that foreign-owned firms were found to be generally larger than domestically-owned firms, with a marginal change in the number of workers found to be significantly associated with a 4.5% increase in the likelihood that the firm was entirely or partially foreign owned.

Noticeable in the empirical studies that have tested the ownership theory is that all are based on developed economies with virtually no evidence from developing countries, SSA in particular. Where similar studies, although scanty, have been conducted in SSA, findings are to some extent dissimilar to those based on developed economies and also to some conducted in developing economies for instance findings by Barthel, Busse and Osei (2008) on Ghana when compared to those by Moss and Ramachandran (2004) on economies in East Africa. Whereas most studies conducted in the developed world, in detail, analyse features of foreign-owned investments, almost none clarifies whether observed superior performance of foreign-owned investments is because such firms are in sectors that are featured by the observed superior characteristics or these foreign-owned firms are simply systematically superior to their domestically-owned counterparts. Empirical studies reviewed and their findings remain inconclusive so far on this matter. For instance, Aaron *et al.* (2014) established that Colombian firms anchored in capital intensive sectors were likely recipients of foreign investment. These findings tend to imply that observed differences between foreign and purely domestically-owned firms are attributable to foreign investors acquiring hitherto domestically-owned firms that were already performing well in terms of capital intensity. This further implies that observed differences may not be due to foreign ownership per se.

In this thesis, I attempt to fill the above identified gaps, manifested in terms of inconclusiveness of findings in the literature and evidence that is still mainly from the developed world. Regarding the later, I provide evidence specific to developing economies, particularly in SSA. Moreover, by employing alternative methods, this thesis provides a robustness check for empirical findings of earlier studies.

Where specific attention has been on those firm-level features amenable to internalisation, firm characteristics like R&D expenditure, advertising, managerial expertise, and marketing skills have received further empirical enquiry, further bolstering the legitimacy of the internalisation theoretical. Can, Dogan & Deger (2017) used Pedroni and Kao Cointegration Tests, as well as Panel Granger Causality methods to investigate the relationship between R&D expenditures, foreign direct investment and economic growth over the period 1996–2011 in a sample of G-7 countries. The results indicated a causality relationship between foreign direct investment and R&D specifically from foreign direct investment to R&D spending at a 1% significance level (Can *et al.*, 2017: 66). Kamata, Sato & Tanaka (2017) used data on Vietnamese manufacturing firms in a study on the internalisation of firms and management practices. Their results indicated that foreign-owned firms had more highly educated managers than domestically-owned firms by a margin of 5%, indicating how skill-intensive foreign direct investments are likely to be when compared to domestically-owned investments. Additionally, 15% more of foreign-owned firms were found to require past managerial experience in similar firms as a recruitment criterion than domestically-owned firms. These findings echoed Saunders's (1982) investigation of the intangible assets hypothesis on Canadian firms. A significant determinant of foreign direct investment activity in his study was management resources. Erdal and Göçer (2015b), who employed panel causality and cointegration methods to investigate the effects of foreign direct investment on R&D and innovation activities in 10 developing economies in Asia, provided further evidence in support of firm level features associated with the internalisation theory. These find a one point increase in the amount of foreign direct investment inflow being associated with a 0.83% increase in R&D expenditures and a 0.42% increase in patent applications in these countries for the 1996–-2013 period (Erdal & Göçer, 2015a: 757). Such findings imply that foreign-owned investments are among other features, associated with high levels of R&D. Empirical findings like those of Kamata *et al.*(2017) imply that differences between foreign and domestically-owned firms in terms of features like management experience are not only identifiable during firm operations and based on years of stay in the industry but also prior to foreign-owned firms taking the decision to invest in the economy in question as well at the time of recruitment. These results may hold even more in case foreign-owned firms target regional markets for exportation of their products. The extent to which a foreign-owned firm supplies regional markets (export oriented) significantly influences the firm's commitment to R&D and hence the associated expenditure (Blomstrom *et al.*, 2000:8).

The more generally accepted and, by consensus, most popular theory, the eclectic paradigm (OLI), has equally received empirical support. Dunning (1988) showed that OLI advantages vary from country to country depending on whether countries are developed or not, whether they are small or large, and on their level and nature of industrialisation (Faeth, 2009: 171). This probably accounts for the variations in earlier empirical tests on the ownership advantages or internalisation theories by studies conducted in developing economies. The OLI theory has been empirically investigated to varying degrees, depending on whether the focus is on location or internalisation advantages, ownership, industries or firms, or even different kinds of foreign direct investment.

In his own empirical assessment, Dunning (1981) investigates whether the competitive advantage of foreign direct investments is dependent on a combination of location and ownership advantages. Dunning's analysis of export and local production data of US manufacturing foreign-owned firms from seven countries finds a significantly negative coefficient of relative market size, and the reverse for skilled employment ratio. More studies testing other ownership advantages have followed that of Dunning's, with some still finding differing results. For instance, Santiago (1987) investigates US firms in Puerto Rico at industry level. Although he finds industry-level foreign investment increasing with relative profits and size of the firms, ownership advantages like capital intensity, relative firm productivity, average firm profits and productivity, are not significant determinants of the share of foreign firms. This dissimilar finding is also evident in other empirical works, for instance in Cai and Guney (2010). In their study of European Union (EU) manufacturing foreign-owned firms in China, using fixed effects modelling and GMM system estimation techniques, results of some ownership advantages like technology and firm profitability do not agree with theoretical predictions, although foreign ownership is still positively associated with export intensity and labour cost.

The significance of ownership and location advantages for foreign-owned firms is affirmed further by Ray (1989) in his investigation on foreign direct investment in the US. At the industry level, Ray finds location aspects, the US growth trend and exchange rate regime, together with ownership advantages, industry R&D intensity, market concentration, capital intensity and industry size significantly explained foreign direct investments. These findings were mirrored by Barrell and Pain (1996), who identified R&D expenditure, the growth and level of gross national product (GNP), firm profits, and the relative production costs in the US compared to other economies positively affected outward foreign direct investment.

In another recent study, Love and Lage-Hidalgo (2010) employ cointegration analysis to empirically investigate the location advantages as set out in the OLI framework. Their study of Mexican foreign direct investments from the US provides further evidence that the market size and the differences in the real wages between the two economies significantly accounted for foreign direct investment flows in Mexico. These results resonate very well with both earlier studies and current empirical studies. For instance, Culem (1988) finds that market size positively influences US foreign direct investment in the European Economic Community (EEC). Martins, Dias & Triches (2017) investigated the determinants of foreign direct investment in Brazil using vector error-correction modelling methods on firm-level data . They find the level of economic activity, wages productivity, the exchange rate regime, and the stability of the national economy to be statistically significant. These results also echo those recently put forward by Tsaurai (2017) in his empirical enquiry into the dynamics of foreign direct investments in BRICS countries. This confirms the explanatory power of the O and L components of the eclectic theory.

Although showing varying findings that suggest further investigation (possibly using alternative methodologies such as cluster analysis), studies based on the OLI theoretical underpinnings have established that combinations of internalisation, location and ownership advantages are the main characteristic determinants of foreign direct investments. Ownership advantages are firm-specific; those with public ownership attributes can be internalised, while the location features are exogenous to the firm. Other factors are more institutional in nature, such as corruption, rule of law and regime type. It is noticeable that empirical evidence guided by the OLI theory does not clearly point out whether foreign-owned establishments systematically differ from domestically-owned firms and so the observed differences in performance may be linked to ownership status of firms. Most studies test Dunning’s theory with greater focus on the location aspects, which does not straightforward answer the empirical question of this study and also leaves the earlier gaps identified. For instance, still most studies focus on western economies like Mexico and the US and utilize conventional methods in their analyses.

Empirical evidence in support of the factor-proportion and proximity-concentration theories is available too in the literature. Several studies have been conducted to this effect although with varying results that further support the need for further empirical enquiry. One such attempt is by Mathä (1999) who uses plant- and firm-level data on Swedish multinationals to investigate whether these theoretical perspectives are supported by data on EU economies. Results from

Mathä's censored regression analyses indicate that horizontal multinationals are promoted if costs of trade are large, firm-level economies are larger than plant-level economies, and R&D intensity is low. Coefficients on economies of scale, factor endowment similarities, host country size, and industry-specific trade costs are found to be significant at either a 5% or a 1% level. The reverse is true for the case of vertical multinationals, supporting the theoretical prediction of the proximity-concentration hypothesis (Mathä, 1999: 15). This result agrees with earlier empirical studies by Brainard (1993a), and Horstman and Markusen (1996).

Mathä further finds the coefficient of R&D is positively significant for vertical relative to horizontal multinationals in his econometric specifications. R&D intensity is found to have more effect on foreign production shares of vertical multinationals, confirming such firms' preference to establish plant affiliates in foreign markets and to export headquarters' services and intermediates to affiliates, a feature of intra-firm trading, as the factor-proportion theory predicts. However, this study finds insignificant effect of factor endowment differences for vertical multinationals, a result that contradicts some earlier empirical findings like Helpman (1984). Helpman uses a general equilibrium model with monopolistic competition in horizontally differentiated goods to explain multinational enterprises (MNEs) but finds that the location of firms abroad is better explained by differences in factor endowments and consequently factor price differences (Faeth, 2009: 175). Additional evidence is provided by Brainard (1993b) who uses US data to examine the extent to which multinational location decisions reveal a trade-off between attaining proximity to customers and concentrating production to realise scale economies. Brainard uses the Probit and OLS regression techniques and finds that, relative to exports, the US's multinational production overseas increases as transport costs rise, as trade and investment barriers decrease, and with lower economies of scale at plant level relative to corporate level. Brainard finds the proximity-concentration theory to be fairly robust (Brainard, 1993b: 538).

Recent empirical evidence, however, indicates that the proximity-concentration hypothesis holds more for non-multi-plant firms in which the trade-off will yield a single option. Otherwise, the coexistence of foreign ownership and exportation is possible and varies with time. Bricongne, Bedoya & Forero (2016) exploit data on French firms to investigate the possible coexistence of foreign ownership and exportation using an analysis based on gravity equations. Their results indicate that substitutability only occurs in core products of the firm, and immediately after the decision to invest in an economy where the demand for the firm's products is high. Empirical studies testing the proximity-concentration theory have limitations similar to those that tested the OLI theory i.e. skewed to the western world, more macro in nature, and do not directly provide answers to our main empirical question.

The knowledge-capital theory has also received empirical attention from various scholarly circles. The evidence, however, has remained inconclusive till to date. Using a panel of inward and outward sales data of US foreign affiliates and some other 36 countries, Carr, Markusen & Maskus (2003) estimate the knowledge-capital model. Their results indicate support for this theory with strong statistical significance found for most variables including; market size, transport costs, and factor endowments. But Blonigen, Davies and Head (2003: 980) contends that Carr *et al.*'s (2003) results are subject to estimation issues, specifically the mis-specification of the underlying theory in their central estimating equation. When one correctly specifies the regression equation, Blonigen *et al.* (2003) argue that the horizontal model cannot be rejected in approval of the KCM. Blonigen *et al.*'s (2003) new estimation results support the horizontal foreign direct investment model against KCM since MNE activity is found to be smaller the more the countries differ in their relative factor endowments. Blonigen *et al.*'s argument might, however, be due to their failure to distinguish between estimation and testing or as a result of complexities surrounding the KCM estimation, given that the relationships predicted by the model are both nonlinear and non-monotonic (Carr *et al.*, 2003: 995).

Mariel, Orbe & Rodr (2007) add to the inconclusiveness of empirical support for the KCM by adopting a time-varying coefficients approach in its estimation. Their OLS estimation of the model based on panel data finds that the vertical component of the KCM is relevant even in the context of European countries with relatively similar factor endowments (Mariel *et al.* 2007: 16). These results echo those earlier found by Braconier *et al.* (2002) in their analysis of Swedish and US outward foreign direct investment data, in which the vertical foreign direct investment was alternatively based on the assumption of a skilled-wage premium as opposed to differences in relative factor endowments. All in all, empirical evidence associated with the KCM remains mixed.

Empirical research on foreign direct investments in relation to domestically-owned investments in Uganda, especially at firm level is still limited, probably owing to the scarcity of firm-level data. Most studies on foreign direct investments in Uganda have been undertaken at macro level, making them inadequate at explaining marginal aspects associated with firm behaviour. Most of such studies have therefore only presented average understanding of the foreign direct investment phenomenon, since macro analyses are informed by aggregates. Some of the studies on foreign direct investments in Uganda have been undertaken by Obwona (1998), Riddervold and Kristiansen (2011), and Wakyereza (2017). As hinted, although they provide good insights on characteristic features of foreign-owned firms in Uganda (particularly Obwona), all these are macro studies. Specifically, Obwona (1998), Wilson and Cacho (2007), and Moss & Vijaya (2004) have produced some likenesses to the international literature. These studies, although still inadequate, find significant differences between foreign-owned and domestically-owned firms in Uganda on various dimensions like; management exp, worker_educ, firm size measured by output, employment levels, and R&D. These studies provide supportive evidence for some of the theoretical frameworks discussed earlier in this chapter.

With reference to the above review, it is clear that empirical investigations regarding foreign and domestically-owned investments have majorly focused on the developed world. Little scholarly attention has been on developing economies, particularly Africa. Even with such bias, findings have been dissimilar in certain cases making it clear that the debate is still inconclusive.

This chapter provides further contribution to the literature especially in Uganda by addressing the issue of the characteristic features of foreign-owned and domestically-owned firms using a relatively novel analytical method that makes use of a rich World Bank Enterprise Survey data set. This chapter attempts to contribute to previous works on Uganda’s foreign and domestically-owned investment enterprises at the firm-level by determining whether the previous findings of the characteristic features and implied differences between foreign-owned and domestically-owned firms still hold or not, when alternative methods of analysis are used. In this endeavour, findings are expected to provide a robustness check for earlier findings, provide evidence that is reflective of developing economies unlike available studies, and ultimately contribute to resolving the inconclusiveness in the literature as highlighted earlier. Given the review of empirical literature, analysis is guided by the OLI and ownership theoretical formulations. Specifically, the latter is micro-oriented and focuses on most of the critical variables that this thesis utilises in the analyses. As will be seen later, analysis in this chapter utilises cluster analysis methods with the help of machine learning techniques to generate clusters of firms, which are summarised and evaluated meaningfully. Choice of cluster analysis as the alternative method was motivated by; (i) its relatively minimal use in economic analyses yet it is a strong tool of multivariate exploratory analysis (Rezankova, 2014). This makes the method suitable given that one of the study aims was to yield results that would act as robustness checks for earlier studies, (ii) its being amenable to machine learning techniques, and (iii) its simplicity in not requiring prior assumptions on the data unlike conventional methods, an attribute that allows this method to straightforward answer the key empirical question at hand. In the following sub-section, a brief description of the principles underlying this method is presented.

2.4 Theoretical Underpinnings of Cluster Analysis

Cluster analysis refers to a set of statistical techniques concerned with exploring data sets to evaluate whether or not they can be summarised meaningfully in terms of a relatively lesser number of groups of objects that are similar to each other; and which are dissimilar in some respects from individuals in other groups (Everitt *et al.*, 2011). It is therefore a numerical technique which entails the classification of multivariate data into a restricted number of groups, which are heterogeneous (dissimilar) between themselves yet homogeneous (similar) within, with the aim of attaining maximum homogeneity and minimum heterogeneity (Johnson & Wichern, 2007).

Historically, cluster analysis can be traced to the period prior to the 2nd World War, originating in anthropology by Driver and Kroeber (1932). The technique was introduced in the field of abnormal and social psychology by Zubin (1938) and Tryon in 1939. It was famously used by Cattell from 1943 for trait theory classification in personality psychology. However, according to Bock (2008), cluster analysis emerged as a major topic in the 1960s and 1970s when the monograph, '*The Principles and Practice of Numerical Taxonomy*' by Sneath and Sokal (1963) motivated world-wide research on clustering methods. The monograph heralded the publication of a broad range of books, such as *Les Bases de la Classification Automatique* (Lerman, 1970), *Mathematical Taxonomy* (Jardine & Sibson, 1971) and *Cluster Analysis for Applications* (Anderberg, 1973). Cluster analysis has witnessed increased use in various disciplines since 2000, with various methods for clustering data having been developed[6]. One of the central elements in cluster analysis is the measurement of proximity of objects under investigation. Proximity, also referred to as association or resemblance of objects, is the logical determination of how similar or dissimilar items or objects under investigation are to each other. Different ways have been developed to measure the proximity of objects being studied

[6] For some of the methods that have been recently developed, see Kaufman and Rousseauw (2005), Rezankova (2009, 2011), and Everitt *et al.* (2011)

but they are generally of two types, namely, the matching type measures and the distance-type measures.

Distance-type measures are specifically employed in the estimation of the distance between objects under investigation, such as firms, individuals or households. They are mostly appropriate for continuous variables. Matching-type measures are usually applied to categorical variables. The use of distance-type measurement is based on a similarity or dissimilarity matrix of *m* by *m* dimension. This matrix contains the pairwise dissimilarities or similarities between the objects under study. For instance, if x_i and x_j are the i^{th} and j^{th} objects respectively, then the entry at the i^{th} row and j^{th} column of the matrix is the similarity s_{ij}, or the dissimilarity d_{ij}, between x_i and x_j.

Within these two types of measuring proximity, there exists a number of techniques for estimating distances between objects or determining the matching between objects. For data containing multivariate continuous variables, the most popular technique is the Euclidean distance estimation or formally the l_2 norm measure[7]. This measure assumes that the space of standardised variables is orthogonal. However, if some components of variables are correlated, weights are used. This measure is specified formally as:

$$d_{ij} = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})\right]^{1/2} \qquad (2.1)$$

where x_{ik} and x_{jk} respectively denote the value of the k^{th} variable of the p-dimensional observations for individuals *i* and *j*. This measure is quite appealing due to its property that d_{ij} can be interpreted as physical distances between two p-dimensional points $X_i' = (x_{i1} \cdot \cdots x_{ip})$ and $X_j' = (x_{j1} \cdot \cdots x_{jp})$ in Euclidean space (Everitt *et al.*, 2011). An observation *i* is declared to be closer or more similar to *j* than to *k* if $d_{ij} < d_{ik}$.

Besides the Euclidean technique, another popular measure is the rectilinear or taxicab measure. This measure is also known as the Manhattan, l_1 norm, or city block distance measure (Everitt *et al.*, 2011). It is generally specified as:

[7] Equally popular are the squared Euclidean distance measure and the city block measure (Tryfos, 1998).

$$d_{ij} = \sum_{k=1}^{p} |x_{ik} - x_{jk}| \tag{2.2}$$

The rectilinear technique measures distances in a street configuration. The Euclidean and rectilinear measures are both special cases of the general *Minkowski distance* or l_r norm measure. Other measures include the Canberra distance, the Pearson correlation, and the Angular separation measures (Everitt *et al*., 2011; Gower & Legendre, 1986). These measures are appropriate when the analyst intends not to use categorical variables in the clustering process.

When the data contains categorical variables, different measures are employed and most of these are defined in terms of the entries in a cross-classification of the counts of mismatches and matches in the "p" variables for two observations. Table 2.2 shows the general version of the cross-classification, while Table 2.3 illustrates the list of similarity measures that can be used.

Table 2. 2: Counts of binary outcomes for two individuals

		Individual i		
	Outcome	1	0	Total
Individual j	1	a	b	a + b
	0	c	d	c + d
	Total	a + c	b + d	p = a + b + c + d

Source: Adapted from Everitt et al. *(2011:46)*

Table 2. 3: Similarity measures for binary data

Measure	Formula
S1: Matching Coefficient	$s_{ij} = (a + b) / [a + b + c + d]$
S2: Jaccard Coefficient (Jaccard, 1908)	$s_{ij} = a / (a + b + c)$
S3: Rogers and Tanimoto (1960)	$s_{ij} = (a + d) / [a + 2(b + c) + d]$
S4: Sneath and Sokal (1973)	$s_{ij} = a\|[a + 2(b + c)]$
S5: Gower and Legendre (1986)	$s_{ij} = (a + d) / \left[a + {}^{1}/_{2}\,(b + c) + d\right]$
S6: Gower and Legendre (1986)	$s_{ij} = a / \left[a + {}^{1}/_{2}\,(b + c)\right]$

Source: Adapted from Everitt et al. (2011:47)

The simple matching coefficient (S1) is the most popular and, as indicated, it is the ratio of the number of matches or mismatches to the total number of attributes. Measures S2–S6 are alternatives developed to counter the shortcomings of S1, for instance, those which arise when dealing with zero-zero matches or uninformative co-absences. For categorical variables with more than two levels, Everitt *et al.* (2011) suggest allocating a score S_{ijk} of zero or one to each variable *k*, depending on whether the two observations *i* and *j* are the same on such variable. An average of the scores over all "p" variables is then generated to give the necessary similarity coefficient specified as:

$$s_{ij} = \frac{1}{p}\sum_{k=1}^{p} s_{ijk} \tag{2.3}$$

However, most of the current data sets contain both categorical and continuous variables; and some of the categorical variables may have more than two levels. This kind of data structure creates a challenge when using the aforementioned proximity measures. One solution is to dichotomise all variables and employ measures for binary data; however, this solution comes at a cost of losing some information due to the transformation. Wright *et al.* (2003) suggests rescaling all the variables so that they are on the same scale. This scaling is done by replacing variable values by their ranks among the observations (objects) and thereafter employing a measure for continuous data. Another option is to construct a dissimilarity measure for each type of variable and combine them, either with or without differential weighting, into a single coefficient (Everitt *et al.*, 2011). Because of the above challenge, more complicated measures for mixed data have been developed. The most popular is the one developed by Gower (1971). Gower's general similarity coefficient caters for mixed data and its varying variable components. Its specification is as seen in (2.4)

$$S_{ijk} = \sum_{k=1}^{p} w_{ijk} s_{ijk} \Big/ \sum_{k=1}^{p} w_{ijk} \tag{2.4}$$

In (2.4), S_{ijk} is the similarity between the i^{th} and j^{th} objects as measured by the k^{th} variable. w_{ijk} is the assigned weight function and typically it is equal to one if the comparison is valid or zero if comparison is invalid. w_{ijk} will also be zero if the outcome of the k^{th} variable is missing for either or both of the objects *i* and *j*. For binary variables and categorical variables with more than two levels, the component similarities, S_{ijk} take the value 1 when the two objects have the same value and 0 otherwise (Everitt *et al.*, 2011). When the variables are continuous, Gower suggested use of the following measure:

$$s_{ij} = 1 - |x_{ik} - x_{jk}| \, / \, R_k \qquad (2.5)$$

where R_k is the range of observations for the k^{th} variable. This means that the city block distance measure is employed after scaling the k^{th} variable to unit range.

After deciding on the proximity measure to use, the next step is to decide on which algorithm to use in the clustering. An algorithm is a step-by-step procedure or method designed to perform an operation, leading to the sought result once correctly performed. Cluster analysis uses algorithms in grouping items and these algorithms are of two categories; hierarchical and non-hierarchical algorithms. In applying hierarchical algorithms, classification involves a series of partitions which may run from a single cluster containing all the observations to "n" clusters each containing a single item (Everitt *et al.*, 2011). Hierarchical clustering can be either divisive or agglomerative. Agglomerative clustering is perhaps the most widely used technique (Everitt *et al.*, 2011). Agglomerative clustering begins with objects considered as individual clusters and the clusters are then stepwise linked until all objects are linked in one cluster (Řezanková, 2014). Divisive clustering follows the reverse procedure by starting with a single cluster consisting of all observations and ending with as many clusters as there are observations. It is the direct opposite of agglomerative clustering (Everitt *et al.*, 2011; Tryfos, 2002). During the process of dividing or agglomerating items into clusters, analysts use several methods at each stage to join or link similar items. These include but are not limited to each of the methods described below:

The single linkage method

This method is also referred to as the MIN version or near neighbourhood method of hierarchical clustering (Tryfos, 2002). Under this method, the proximity of two clusters is defined as the minimum of the distance between any two points in two different clusters. The minimum distance is actually a measure of maximum similarity. In this method, only pairs consisting of one individual from each group are considered (Everitt *et al.*, 2011: 74).

The complete linkage method

This method is also referred to as the MAX version of hierarchical clustering or the furthest neighbourhood method (Everitt *et al.*, 2011; Tryfos, 2002). It is the direct opposite of single linkage method and, in this case, the distance between two clusters is equivalent to the distance between their two most distant members.

The average linkage method

The single linkage and complete linkage are not without their own shortcomings. A compromise between these two is the average linkage method. Under this method, the distance between any two clusters is defined as the average of the distances of all pairs of observations, with one observation in the pair taken from the first cluster and the other observation taken from the second cluster (Tryfos, 2002). According to Everitt *et al.* (2011: 76), it is also referred to as the unweighted pair group method (UPGMA), which uses the average approach.

The centroid method

Also referred to as the unweighted pair group method, the centroid method derives its name from the use of the centroid approach. It entails merging clusters that have the most similar mean vectors and instead of a proximity matrix it uses a data matrix (Everitt *et al.*, 2011). The method involves computing the geometric center (centroid) of each cluster. The distance between any two clusters is then defined as the distance between the two centroids.

Median linkage

This method is also known as the weighted pair group method (WPGMC) using the centroid approach. In this method, the centroids of the constituent clusters are given equal weights to produce the new centroid of the merged cluster. This method reduces the likelihood of objects in the most frequent of the pair of clusters to be merged, dominating those in the smaller cluster. It therefore yields an intermediate new centroid between two constituent clusters.

Ward's method

Ward introduced this method in 1963. It involves merging clusters based on the size of an error sum-of-squares criterion. The key aim at each stage of cluster merging is to minimise the increase in the total within-cluster error sum of squares. The error sum of squares (E) is given by:

$$E = \sum_{m=1}^{g} E_m \tag{2.6}$$

where

$$E_m = \sum_{l=1}^{n_m} \sum_{k=1}^{p_k} \left(x_{m_{l,k}} - \bar{x}_{m,k}\right)^2 \tag{2.7}$$

in which $\bar{x}_{m,k}$ is defined as $(1/n_m) \sum_{l=1}^{n_m} x_{m_{l,k}}$ which is the mean of the m^{th} cluster for the k^{th} variable and $x_{m_{l,k}}$ is the score on the k^{th} variable for the l^{th} object in the m^{th} cluster. This method utilises Euclidean distances and is mainly applicable to continuous data.

The non-hierarchical Clustering algorithm or method is popularly referred to as the partitioning method. Just as with the hierarchical clustering method, there exist several ways of cluster formation under the partitioning approach. One of the most popular methods is the K-means procedure. Unlike the previous methods, which are based on the Euclidean distances or city block distances, the K-means method uses the within-cluster variation as a measure to generate homogeneous clusters (Mooi & Marko, 2011). Its fundamental objective is segmenting data in such a way so as to minimise within-cluster variation as much as possible.

Tryfos (2002) describes the K-means method as consisting of the following steps;

1. Step 1: Specify the number of clusters and, arbitrarily or deliberately, the members of each cluster.

2. Step 2: Calculate each cluster's centroid and the distances between each observation and the centroid. If an observation is nearer the centroid of a cluster other than the one to which it currently belongs, re-assign it to the nearer cluster.

3. Step 3: Repeat Step 2 until all observations are nearest the centroid of the cluster to which they belong.

4. Step 4: If the number of clusters cannot be specified with confidence in advance, repeat Steps 1 to 3 with a different number of clusters; and conduct an evaluation of the results.

In terms of preference by most analysts, hierarchical agglomerative clustering methods have taken the lead in so far as most empirical studies are concerned. One of the explanations has been that the creation of a taxonomy generally requires a hierarchy. Elsewhere, this method has been associated with the ability to generate high-quality clusters. However, even with such strengths, there is no consensus so far as to which algorithm is the most appropriate in any field of study. This lack of consensus is because any algorithm will find clusters in a data set, even when such a data set does not naturally have the clusters. Indeed, there is no clustering function "f", for each n >= 2, which satisfies scale invariance, richness and consistency (Kleinberg, 2002). Because of this reality, several scholars have devised cluster evaluation methods and robustness checks in the field of cluster analysis. For instance, after a particular algorithm has been applied in clustering data, cluster validity can be performed by use of the cohesion and isolation methods. This evaluation and robustness check can further be coupled with a rerun of the analysis process on the same data set using different algorithms or methods (Everitt *et al.*, 2011; Gallegos & Ritter, 2005; Mooi & Marko, 2011). Further, assessments of validity of results can be conducted based on the extent to which clusters are parsimonious, accessible, and actionable, among other criteria (Naughtin & Rankin, 2016). Given the merits associated with cluster analysis as earlier mentioned, numerous scholars have utilised it in both natural and social science research, let alone in business and marketing. As the following brief review of literature confirms, clustering techniques have been embraced in many empirical fields, except in economics where it has been not widely utilized.

2.5 Cluster Analysis: A brief review of empirical Literature

A significant body of literature regarding the empirical application of cluster analysis does exist and continues to grow. Its existence is visible in various fields such as marketing, medicine, psychology, development, socioeconomic field, science and innovation, and taxation. In marketing and business, cluster analysis has been used to study and consequently segment clients for the purpose of developing and marketing profitable goods and services. By clustering consumers based on similar attributes, segments can be created, promotional messages promptly targeted, and products customised to such segments. James and Michael (2008), using a household sample of 500 grocery shoppers, employed cluster analysis techniques to segment shoppers and to understand their characteristics based on the preferred characteristics of their grocery store choices. This study used a K-means clustering algorithm. They found that opening times and competitive prices were the most influential factors in the choice of grocery store for Alabama shoppers. Scholars, such as Mangaraj and Senauer (2001), Lynn (2011) and Fânaru (2016), have conducted similar studies using cluster analysis in marketing.

In medicine, scholars such as Beckstead (2002) have used cluster analysis to investigate questions related to the field of nursing, while Clinton *et al.* (2004) employed cluster analysis to investigate key diagnostic variables influencing eating disorders among patients. Clinton's study used ten key clinical variables of primary significance in diagnosing eating disorders among patients. Vogt and Nagel (1992) highlights the strength of cluster analysis in clinical diagnosis while Liao, Li, Kianifard, Obi & Arcona (2016) applied K-means and hierarchical cluster analysis with various linkage methods onto health claims data to investigate end-stage renal disease patients who initiated haemodialysis. In astronomy, cluster analysis has been utilised in organising star catalogues, for instance the second Palomar Observatory Sky Survey (POSSII), in Yoo *et al.*, (1996). In psychology, Borgen and Barnett (1987) articulated the indispensability of cluster analysis in counselling. Ullrich-French and Cox (2009), applied cluster analysis to explore naturally occurring combinations of regulation motivations in the US among physical education students. The study identified motivation profiles of students based on the four types of motivation i.e. intrinsic, identified, external and introjected motivation. Using both K-means and hierarchical clustering approaches, Ullrich-French and Cox (2009), confirmed that the relative levels of students' self-determination may not distinguish among physical education experiences for those students with certain combinations of motivation regulations. Elsewhere, Tanton, Dodd, Woodfield & Mabhala (2015) used two-step cluster analysis to study the eating behaviours of British university students.

As much as the use of cluster analysis in multiple disciplines commands significant popularity, similar application in other social sciences gained significance only recently, especially in the field of economics. Surprisingly, the roots of the clustering concept are highly associated with early works of Marshal (1920) who identified a triad of external economies that tend to account for local clustering of economic activity (Lazzeretti, Sedita & Caloffi, 2012). The refining of the concept in the 1990s has mainly been attributed to economic geographers and exponents of New Economic Geography (NEG), especially the works of Paul Krugman (1991). The use of cluster analysis, however, has now gained popularity among economic scholars and policy makers at different levels of engagement (Lazzeretti *et al.*, 2012; Murry, 2016). There is clear evidence regarding the growing use of cluster analysis in development economics in a bid to stimulate and direct transformational policies. Ieva Brauksa (2013) used cluster analysis to study the differences in municipal levels of development in Latvia and the factors, which account for the observed differences. By classifying municipalities based on economic indicators of development, as opposed to only geographical proximity, the study identifies municipalities in urgent need of development stimulation. At a global level, Grein *et al.*(2010) uses hierarchical clustering techniques to study the role of corruption and the implications for global firms using a sample of 39 countries. Other than grouping countries based on geographical proximity, emerging versus developed economies, or middle versus low income perspectives, clustering allowed Grein *et al.* (2010) to cluster countries based on economic, technological, cultural and quality of life covariates. Grein *et al.* (2010) sheds light on how corruption has altered the conventional market analysis which global firms have to undertake in the process of choosing where to invest, and related strategy formulation. Grein *et al.*'s (2010) study opens gates for further research focused on hypothesis testing involving corruption and other key variables that influence the economic environment in which global firms operate.

Cluster analysis has also been applied in the field of socioeconomic research with numerous studies about this. Lombardo and Falcone (2011) uses partitioning around medoids, a clustering method, to investigate the interactions between criminal activity, geographical location, and economic performance. Among other findings, the study provides evidence regarding significant disparities in the spatiotemporal evolution of crime across provincial units in Italy. Additionally, criminal activities are found to be not inseparably associated with geographical location, as it had been previously proposed in scholarly literature. Findings of Lombardo and Falcone (2011) attest to the ability and merits of cluster analysis in providing robustness checks for earlier empirical findings, especially where other conventional methods were applied. Elsewhere, Rovan and Sambt (2003) investigated socioeconomic differences between Slovenian municipalities using the hierarchical and K-means clustering approaches for the first and second phases of their investigation. This study aimed to provide development policy evidence regarding the need to keep differences in economic indicators among regions in sustainable limits for the general welfare of the economy in question. Study findings were in agreement with studies that attempted to identify and characterise welfare regime typologies, for instance in Minas *et al.* (2014). Cluster analysis has also been used in the field of science and innovation in the scholarly works of Yurtseven and Tandogan (2012) and Hollenstein (2003). In taxation literature, cluster analysis has been applied in predicting business firms with a high likelihood of engaging in tax evasion and related fraudulent behaviour. This is evident in the works of scholars such as Dias *et al.* (2016) and Nikola (2015), whose findings are very informative around policy, especially those targeting foreign direct investments.[8]

Studies on foreign direct investments have also recently made use of cluster analysis in attempts to either answer new research questions or to confirm earlier findings. Most of these studies, however, have not specifically delved into firm-level characteristics and the issues relating to the investment determinants of foreign-owned in relation to domestically-owned firms, as this study does. One such a study is by Simionescu (2016) who used cluster analysis, Bayesian techniques and panel data approaches to investigate the relationship between foreign direct investments and economic growth in 28 countries of the EU during the recent economic crisis. Other studies that employed cluster analysis on foreign direct investment and growth include; Forte and Santos (2015), Gutiérrez Portilla *et al.* (2016), and Yu & Zhang (2007). As it may be noted, most of these studies employed clustering at either macro or international levels, with fewer or none at micro firm level and focusing on the area of this thesis.

In Uganda, literature in the field of economics where cluster analysis techniques have been applied is either unavailable or minimal. Cluster analytical techniques have been used primarily in psychology, medical, biological, and agricultural sciences. Namisango *et al.* (2015) used cluster analysis to study symptoms of ambulatory HIV/AIDS patients in Uganda. Ward's method of linkage was used to link patients based on symptom occurrence and a five-cluster solution was achieved (Namisango *et al.*, 2015). Yada *et al.* (2010) used cluster analysis to characterise Ugandan sweet potato germplasm in which the use of fluorescent-labelled simple sequence repeat markers was made. The study sought to characterise the diversity of selected but superior sweet potato genotypes for purposes of conservation and breeding. Díaz-Bonilla and Thomas (2016) undertook a related study in which hierarchical and K-means clustering methods were applied to develop country typologies of food security. Clustering enabled the investigators to identify countries that were more food insecure relative to others. It is evident therefore that the use of cluster analysis in economics and particularly in Uganda is extremely scarce, more so at firm-level analysis. Its application to foreign and domestic firms' analysis

[8] For more on application of cluster analysis in the taxation field, see Rybová (2015), Mihóková *et al.* (2016), and Andrejovská and Hudáková (2016).

will immensely enrich scholarly understanding of how these enterprises differ and behave in terms of performance.

The use of cluster analysis in this study differs to some extent from how the cluster concept is applied in the field of industrial economics. In industrial economies, clusters are used to refer to geographic concentrations of industries related by inputs, skills, demand, knowledge, and other useful industrially-related linkages (Delgado, Porter & Stern, 2014: 1787; Kuah, 2002: 207). Clusters and their use in industrial economics are more linked or equated to agglomeration in which prior assumptions are postulated before studying industrial groups. In industrial economics, although agglomeration has been recognised for generations, the cluster concept is more associated with earlier works on nations' competitive advantages; as per the work of economist Porter who used concept of clusters in the early 1990s. Porter defined clusters as groups of interconnected firms, suppliers, related industries and specialised institutions in specific fields, visible in specific locations. Recently, however, clustering algorithms similar to those applied in this thesis have also been explored to assess and study groups of closely related industries (Delgado *et al.*, 2014: 1791). Therefore, whereas industrial economists use clusters to investigate value chains and industrial characteristics associated with geographic concentrations of industrial establishments, this chapter differs to some extent. This chapter uses clustering techniques that are exploratory in nature without imposing assumptions *a priori*. Using Ugandan data, this study generates groups of firms that are homogeneous within but distinct between themselves; it analyses the characteristics of these groups; and it compares the results to earlier studies and conclusions, which underpin literature.

This chapter will greatly contribute to the growing literature in many ways. It is one of the first studies to utilise unsupervised, machine-learning techniques, specifically cluster analysis, to analyse private firm enterprises in Uganda. Additionally, while some studies have applied cluster analysis in Uganda, this application has been mainly in natural sciences; and to the best of the authors' knowledge, this study is one of the first studies to use this type of analysis in the economics field on Ugandan data. Cluster analysis presents a methodological opportunity to generate results to either confirm or dispel earlier empirical findings. The next section presents a description of the methodology, which was used in this chapter.

2.6 Methodology and Results

The overall analytical aim of this chapter is to examine the characteristics of foreign-owned and domestically-owned enterprises in Uganda; and to gain empirical insights into how they differ and whether observed performance differences arise from comparisons between different sub-groups or whether foreign-owned firms differ systematically across a number of dimensions. I also aimed to establish whether a firm's ownership status, i.e. foreign or domestic, is an important factor that helps, with other variables, to group particular firms into certain clusters. I adopted a three-stage methodology, using World Bank survey data on firms in Uganda.

2.6.1 Data Description and Transformation

In this chapter, I used data from the World Bank Enterprise Surveys (ES). The ES collects data from key manufacturing and service sectors in every region of the world. Sampling for the ES data collection is based on stratified random sampling. In this case all population units (in this case firms) are clustered within homogeneous clusters and simple random samples are selected within each cluster. This method enables computing estimates for each of the strata with a specified level of precision. This inherent precision has implications regarding the validity of results obtained from these data, given the methods employed. The sampling equally involves weighting which caters for the varying probabilities of selection across different strata. This weighting implies that population estimates can equally be obtained with a certain level of precision. In my clustering, I utilise linkage methods that incorporate weights in the analysis so as to lend credence to the results from the context of all private enterprises in Uganda.

The ES use standardised survey instruments and a uniform sampling methodology to minimise measurement error and to yield data that is comparable across the world's economies. This implies that results obtained from analysing Uganda can also be comparable to other relatively similar economies around the world. The primary sampling unit of the surveys is the establishment or firm, being a physical location where industrial production or provision of services takes place. Large firms are intentionally oversampled because most economies have mainly small and medium enterprises. This might have implications for the results from the analyses and hence the need for some caution while utilizing them and attaching meaning.

The surveys cover 13 topics, being: control information; general information; infrastructure and services; sales and supplies; degree of competition; capacity; land; crime; finance; business-government relations; employment; business environment; and performance.

For Uganda, the surveys have been conducted more than once; and for this study, I used the most recent, being the 2006 and 2013 surveys. The firms are industrially classified into the following: food; textiles; garments; leather; wood; paper; publishing, printing and recorded media; chemicals; plastics and rubber; non-metallic mineral products; basic metals; fabricated metal products; machinery and equipment; electronics ; transport machines; furniture; recycling; construction; motor vehicle servicing; wholesale; retail; hotels and restaurants; transport; and information technology.

2.6.1.1 Variable Selection, Measurement, and Transformation

Selection of variables used in the analysis was guided by; the theoretical analyses earlier discussed, most especially the ownership advantages and eclectic theory, earlier studies on Uganda by scholars such as Obwona (1998) and Wilson and Cacho (2007), studies related to the topic but conducted elsewhere, and subject matter knowledge. Scholars who have analysed firms in other economies on related topics have also used many of these variables in their analyses. A case in point on the South African Economy are the works by Rankin (2013), and Naughtin & Rankin (2016). Moreover, some scholars have it that in clustering, common sense also plays a role, as the number of legs can be used to differentiate spiders from insects but may not be helpful in differentiating between spiders and cats. I selected variables thought to be relatively associated with firms' performance. These variables are presented in Table 2.4, which includes their labels and how they are measured. However, I generated other variables when deemed necessary during the analysis.

Table 2. 4: Variable description

Variable name	Variable label	Variable Measurement
Employment	ln_employment	log (№ of fulltime employees)
Labour productivity	ln_lbr_prdvty	log (real firm output/ № of employees)
Foreign ownership	fdi-stake	% of firm shares owned by foreigners
Capital intensity	ln_capital_intensity	log (NBV machinery/№ of employees)
Material per worker	ln_material_perworker	log (real value of materials/ № of employees)
Management experience	ln_mgt_exp	log (№ of years spent in industry)
Wage	ln_wage	log (total labour cost/ № of employees)
Worker education	ln_worker_educ	log (№ of years of employee education)
Absolute firm exports	ln_exports_absolute	log (%of exports*real firm sales)
Firm age	firm_age	year of survey minus year of firm start
Firm output	real_sales	log (real firm sales)

Source: Author's own description based on data

I performed a log transformation of the variables in order to remove any inherent skewness and to make patterns in the data more interpretable. After this transformation, descriptive statistics were generated. Variables used in clustering further underwent a linear transformation. These variables were standardised so that all have a mean value of 0 and a variance of 1. This standardisation was performed because the data consists of variables measured using different scales which made comparisons challenging. Also, centring variables makes them more compatible with cluster analysis. I divided the data set into three subsets, namely: firms surveyed in 2006; firms surveyed in 2013; and a combination of the two surveys. I named these in the following order: NData2006, NData2013 and CNData. The combined data set would provide a much larger sample for the cluster analysis while further analysis of each survey specifically was thought to shed light on the consistency of the clustering algorithm and hence lend credence to the results. This is what motivated the creation of the three sub-data sets. Because of this division and also due to the need to use values devoid of the inflation effect in the analysis, I deflated some variables for both NData2006 and NData2013 before combining the two subsets to form CNData. I thus used the 2006 and 2013 GDP deflators and obtained real values based on the 2002 base year, as provided by the Ugandan statistics bureau. In this analysis, I specifically defined a foreign-owned firm as one with a minimum of 10% ownership shares belonging to individuals or parties from nationalities other than Ugandan. Choice of the 10% threshold was motivated by the OECD's (2008) benchmark definition of foreign direct investment. Whereas a 50% threshold has been used by some scholars and would probably be more suitable for an economy like Uganda, the 10% threshold was thought to provide more space for the algorithm to group firms based on similarities at the lowest level of foreign ownership stake. This would add to the nuance of the comparisons. In the descriptive statistical analysis and regression, the fdi-stake is a dummy variable equal to 1 for foreign-owned firms and 0 otherwise. However, in the clustering stage of the analysis, fdi-stake is used as a continuous variable.

2.6.1.2 Dealing with missing values

Missing values for some or all variables of interest is one of the biggest challenges to clustering data. A missing value, represented by NA in statistical software packages like R, is a placeholder for a datum of which the type is known but its value is not. Therefore, it is difficult to perform statistical analysis on data where one or more values in the data are missing (de Jonge & van der Loo, 2013). The survey data I used in this study also had missing values on several variables. There are several ways of dealing with missing values when undertaking cluster analysis. These include partial data cluster analysis, replacing missing values with means and imputation, all of which are discussed below.

2.6.1.3 Partial data cluster analysis

This method involves grouping items based on the data that they have in common. Whereas it improves on other methods like complete case clustering, it is not the recommended method for dealing with missing values. Partial analysis may leave out important information in the process.

2.6.1.4 Replacing missing values with means

Another method of dealing with missing data is to replace missing values with the average values of the variables in question. Such mean replacement, however, fundamentally alters the underlying structure of the data. The implication of this replacement method is that when clustering is performed, the resulting clusters are to some extent a consequence of the decision taken by the analyst to replace missing values, with means as opposed to the data itself determining fully the clusters (Bock, 2017).

2.6.1.5 Imputation

This method involves replacing missing data with sensible estimates of the values missing. It is based on a similar logic to that which underlies partial data cluster analysis. It is, however, more superior to partial data analysis because it takes into account relevant information that is usually ignored by partial cluster analysis. However, it requires the use of modern imputation algorithms, mainly those integrated in R like those in multivariate imputation by chained equation (MICE) and multiple imputation (MI) packages (Bock, 2017)[9]. Using imputation ought to be done skilfully. It is advisable to use it concurrently with partial data cluster analysis for comparison purposes.

In this chapter, I used the imputation method to deal with missing values. There are several techniques for employing the imputation method. These include the basic numeric imputation modelling, hot deck imputation techniques and kNN-imputation[10]. I also used the hot deck imputation technique, using the random procedure. In hot deck imputation, missing values are filled in by copying values from similar records in the same data set. By notation, we let $\widehat{\Psi}_i = \Psi_j$ where Ψ_j is taken from the observed values. $\widehat{\Psi}_i$ is the imputed value for a missing value for variable Ψ. Moreover, hot deck imputation is applicable to both numerical and categorical data as long as enough donor records are available (de Jonge & van der Loo, 2013). However, I imputed values for missing data only when clustering because the algorithms I used are not compatible with missing values i.e. NAs. After clusters have been formed, I describe them based on original data-set variables.

2.6.2 Study analysis

In my attempt to understand the characteristics of foreign-owned and domestically-owned firms, whether the observed differences between foreign-owned and domestically-owned firms are due to comparisons between different sub-groups, or whether foreign-owned firms and domestically-owned firms differ systematically across a number of performance dimensions, I conducted two stages of analysis.

The first stage of the analysis involved descriptive statistics, using means and medians for numeric variables and frequencies for non-numeric variables. I did this separately for foreign and domestic firms to compare the two groups. This descriptive analysis allowed me to

---

[9] Bock is the founder of Diplayr, an R package that is useful in cluster analysis.

[10] For more information, see Edwin & Mark's Introduction *to Data Cleaning with R (2013).*

investigate properties of the distributions of the variables, such as the normality property. This first stage also involved the use of regression analysis, using firms' ownership status (fdi-stake) as the regressor and selected variables as regressands and controls. Some of the regressors used include employment, labour productivity, management experience and wage. Using firm ownership status as a dummy equal to 1 if a firm is foreign-owned and 0 otherwise, this analysis allowed me to investigate how foreign ownership is likely to be associated with aspects like firm size, productivity, and wages. I ran these regressions using an R package called 'robust', found in the CRAN repository. The subsection that follows presents the results of the descriptive statistics and regression.

2.6.2.1 Descriptive Statistical Results

In Table 2.5, the weighted descriptive statistics in terms of mean and median real values for the numerical variables are presented. These were generated based on the overall data set. The first column contains values for firms in the overall data set, while the other two columns contain values for foreign and domestic firms respectively.

Table 2. 5: Descriptive statistics, weighted mean and median (real) values CNData-combined data set

Variable	All (N=1319)	Foreign(N=202)	Domestic(N=1117)
Mean			
log employment	2.41	3.07	2.31
log labour productivity	15.68	17.14	15.46
log capital intensity	14.14	14.92	13.96
log material/worker	14.46	15.74	14.26
log management experience	2.16	2.33	2.13
log wage	13.53	14.07	13.44
log workers' education	2.43	2.44	2.43
log exports	18.80	20.93	18.12
actual employment	23.86	42.20	21.02
Median			
log employment	2.20	3.00	2.08
log labour productivity	15.48	16.67	15.34
log capital intensity	14.16	15.32	13.85
log material/worker	14.33	15.88	14.23
log management experience	2.30	2.40	2.20
log wage	13.63	13.96	13.58
log workers' education	2.56	2.64	2.56
log exports	18.16	21.06	17.96
actual employment	9.00	20.00	8.00

Source: Author's own computations based on World Bank Enterprise Survey, 2006/13 panel

With reference to Table 2.5, it is observed that overall mean and median values for all variables, except actual employment, are approximately similar. This approximate similarity is suggestive of a symmetric distribution exhibited by the variables. Actual employment is an outlier because it was not subjected to log transformation. It is further evident that the mean and median real values for the overall column and the column for domestically-owned firms are approximately similar. Nevertheless, big differences are observed between the values of the second column, foreign firms, and the other two columns. The only exception is worker_educ values, which are almost uniform across all columns. These statistics seem to suggest that significant differences between firms in the data set are more associated with firm ownership status i.e. foreign versus domestic ownership than with other aspects. This is one of the key hypothetical assertions I sought to validate using cluster analysis.

Table 2.6, shows descriptive statistics in the form of weighted percentage frequencies for non-numeric variables. These were generated based on overall data and the NData2013. I eliminated NData2006 due to the high level of missing values.

In Table 2.6, the first two columns indicate the weighted percentage composition of foreign-owned and domestically-owned firms respectively, computed based on the overall data set (CNData). The last two columns indicate the weighted percentage compositions of foreign and domestically-owned firms respectively, computed based on specifically foreign and domestic subsets for firms surveyed in 2013. Because computation was based on these specific sub sets, the frequencies are reported by column for easy intra-firm comparison[11]. It is noticeable that in the first two columns, domestically-owned firms' percentages outweigh foreign-owned firms' percentages. This difference is probably due to lower numbers of foreign-owned firms relative to domestically-owned firms. Looking at the last two columns of 2013 surveyed firms, intra-percentage comparisons on the sector level indicates that foreign-owned firms are more likely to be in either the "others" category (54.7%) or manufacturing (26.4%) than retail. This feature is the reverse for domestically-owned firms where actually the lowest percentage (25.5%) of these firms is in manufacturing. The general picture for both types of firms reveals that the largest percentage is not in manufacturing. Less presence in the manufacturing sector has serious implications for job creation. This might be an initial indicator to the reasons for the earlier mentioned dismal economic outcomes even when foreign investments continue to flow into the country. Additionally, over 80% of foreign-owned firms in the 2013 sample were located in Kampala, compared to a similar 74.1% of domestically-owned firms. This urban concentration around Kampala of both foreign and domestically-owned firms is probably due to easy access to markets and better infrastructure being available in urban areas. There are also many benefits associated with agglomeration. In terms of firm size, as measured by the number of workers, foreign-owned firms have higher values for large firms at 6.6% and for medium firms at 30.8% when compared to domestically-owned firms at 1.2% for large firms and 11.6% for medium firms. This implies, as revealed in the literature reviewed, that foreign-

[11] The overall sample allows comparison between foreign and domestic firms, while the 2013 sample was used to do comparisons within the categories (foreign/domestic) themselves.

owned firms are more likely to be larger in size when compared to domestically-owned firms. The statistics also reveal that generally most firms are either small or medium in size whether domestic or foreign-owned in Uganda.

Table 2. 6: Descriptive statistics for qualitative variables, weighted (%) frequency

Variable	Overall sample		2013 sample	
Sector type	Foreign (%)	Domestic (%)	Foreign (%)	Domestic (%)
Manufacturing	10.7	89.0	26.4	25.5
Retail	6.2	93.5	19.0	33.2
Others	13.3	86.6	54.7	42.3
Location				
Kampala	11.7	88.3	84.1	74.1
Jinja	14.8	85.2	7.8	5.2
Mbale	1.7	97.4	0.8	5.2
Lira	2.7	97.3	0.6	2.4
Mbarara	8.8	90.4	3.2	3.8
Wakiso	4.3	95.4	3.6	9.3
Firm size (№ of workers)				
Large (>=100)	37.9	59.0	6.6	1.2
Small (>=5&<=19)	7.7	92.2	62.6	87.2
Medium (>=20&<=99)	23.4	76.1	30.8	11.6
Firm age				
<10	8.7	91.2	54.5	55.1
>=10&<=50	8.8	91.0	44.4	44.7
>50&<90	32.6	58.3	1.10	0.20
R&D Expenditure				
No	7.5	92.4	59.3	71.4
Yes	12.2	87.4	40.7	28.6
Formal employee training				
No	9.7	90.2	60.9	65.9
Yes	11.7	88.0	39.1	34.1

Source: Author's own computations based on World Bank Enterprise Survey, 2006/13 panel

Many firms in Uganda, as the results in Table 2.6 suggest, are not likely to prioritise R&D expenditure and formal training of employees. The percentage of firms that prioritise R&D expenditure is only 40.7% and 28.6% for respectively foreign and domestically-owned firms. R&D spending has been found to greatly influence firm performance and is likely to be more prioritised when such firms pursue product differentiation strategies. The results therefore might imply that domestic and foreign-owned firms tend to pursue different strategies, which ultimately translates into performance differences between the two types of firms. Most foreign firms engage in product differentiation so as to gain a competitive edge in the host economy. The percentage of firms that provide formal employee training is almost half as much as those that don't provide both for foreign-owned (39.1% relative to 60.1%) and domestically-owned firms (34.1% relative to 65.9%). These descriptive statistical results are a preliminary pointer to the hypothesized differences between foreign-owned and domestically-owned firms in Uganda, a hypothesis I ultimately test using clustering methods. While these statistics may point to the assumed differences, these might be incidental and not relational in nature. To further understand the hypothesized differences, I employed regression methods in an attempt to establish existence/absence of association between firms' ownership status and the performance variables considered in the analysis. Procedures and results of this stage of analysis follow;

2.6.2.2 Regression Results

Using overall data (CNData), I estimated robust regressions using four dependent variables, namely employment as a proxy for firm size, labour productivity, wage, and management experience. In each case, a dummy variable for foreign-ownership status was used as a regressor to investigate if and how foreign ownership was likely to be associated with the dependent variables. I also used other variables as controls in order to gain more insights. Additionally, I broke foreign ownership into categories and regressed them against the dependent variables mentioned. In this attempt, I aimed to gain insights into whether categorical levels of foreign ownership are likely to differentially be associated with variations in the dependent variables. In Tables 2.7 and in 2G in the appendices, M1, M2 and M3 correspond with the three regression equations that were estimated for each dependent variable. In M1, I used only the dummy variable as the regressor; in M2 I used fdi categories as the regressors; and, in M3, I again used the fdi dummy, this time with controls.

Table 2. 7: Regression results of firm ownership on employment and labour productivity

Variables	log employment			log labour productivity		
	M1	M2	M3	M1	M2	M3
fdi dummy	0.75***		0.66***	1.57***		2.09***
	(0.019)		(0.024)	(0.055)		(0.101)
fdi dummy(10% - 40%)		1.213***			1.87***	
		(0.102)			(0.26)	
fdi dummy (41-60%)		1.30***			2.17***	
		(0.069)			(0.184)	
fdi dummy (61-99%)		0.89***			2.96***	
		(0.052)			(0.134)	
all foreign		0.69***			1.34***	
		(0.022)			(0.064)	
sector type-retail			-0.30***			0.02
			(0.019)			(0.081)
sector type-others			-0.15***			0.38***
			(0.018)			(0.074)
location cat-kampala			0.16***			1.03***
			(0.031)			(0.139)
location cat-lira			-0.22***			-0.25
			(0.057)			(0.209)
location cat-mbale			-0.24***			0.68***
			(0.045)			(0.199)
location catmbarara			0.28***			0.90***
			(0.049)			(0.188)
location catwakiso			-0.13***			0.45***
			(0.039)			(0.180)
Constant	2.500***	2.511***	2.57***	15.52***	15.50***	15.26***
	(0.032)	(0.031)	(0.061)	(0.058)	(0.057)	(0.132)
Observations	1297	1296	734	1054	1053	493
RSE	0.630	0.630	0.603	1.57	1.58	1.57

Note: **p<0.1; **p<0.05; ***p<0.01*

figures in parenthesis are standard errors

Source: Author's own results generated based on World Bank Survey Data 2006/13 panel

In Table 2.7, it is evident that firms that are foreign-owned are more likely to be associated with higher employment levels, and hence likely to be larger when compared to their domestic counterparts. They are also likely to be associated with higher levels of labour productivity. In M1, for instance, it is evident that foreign-owned firms tend to have 75% more workers (fdi dummy = 0 used as a reference) when compared to domestically-owned firms. This result is significant at 1%. In M2, fdi dummy categorisation reveals similar results, with the 40-60% category having the largest coefficient. M3 shows that, even after controlling for location and sector type where firms are engaged, firm size in terms of employment numbers remains highly associated with ownership status. The fdi dummy coefficient of 0.66 is significant at 1%. For labour productivity, coefficients are not only significant but higher relative to employment. Results still indicate that foreign-owned firms are more likely to be associated with higher levels of labour productivity. In Table 2G in the appendices, when wage and management experience are used as the regressors, then it is again noted that coefficients for fdi dummy are significant at 1%, though they are lower than those in Table 2.7. Specifically, taking fdi dummy = 0 as a reference, foreign-owned firms tend to have 54% more in wages when compared to domestically-owned firms. This result is significant at 1%.

The regression results in Tables 2.7 and 2G are further supported by illustrative figures. Figures 2.4, 2.5 and 2.6 are box plots of firm ownership status versus; labour productivity, management experience, and average wage, all derived from the overall data set. In the box plots, the horizontal line in the middle of each box corresponds to the median value of the variable in question for both foreign and domestically-owned firms. The end of the upper whisker corresponds to the maximum value while the end of the lower whisker corresponds to the lowest value. The dots at the end of the whiskers indicate outliers. Focusing on Figure 2.4 first, it is noticeable that the lowest value of labour productivity for domestically-owned firms is about 11 with several outliers while for foreign-owned firms the minimum is exactly 12 without any outliers. Foreign-owned firms post a maximum log value of about 21 while domestic firms post a maximum of about 19. More than half of foreign-owned firms post productivity values above 16 unlike domestically-owned firms whose median level is even below 16. What is observed in Figure 2.1 is to a large extent reflective of figures 2.2 and 2.3.

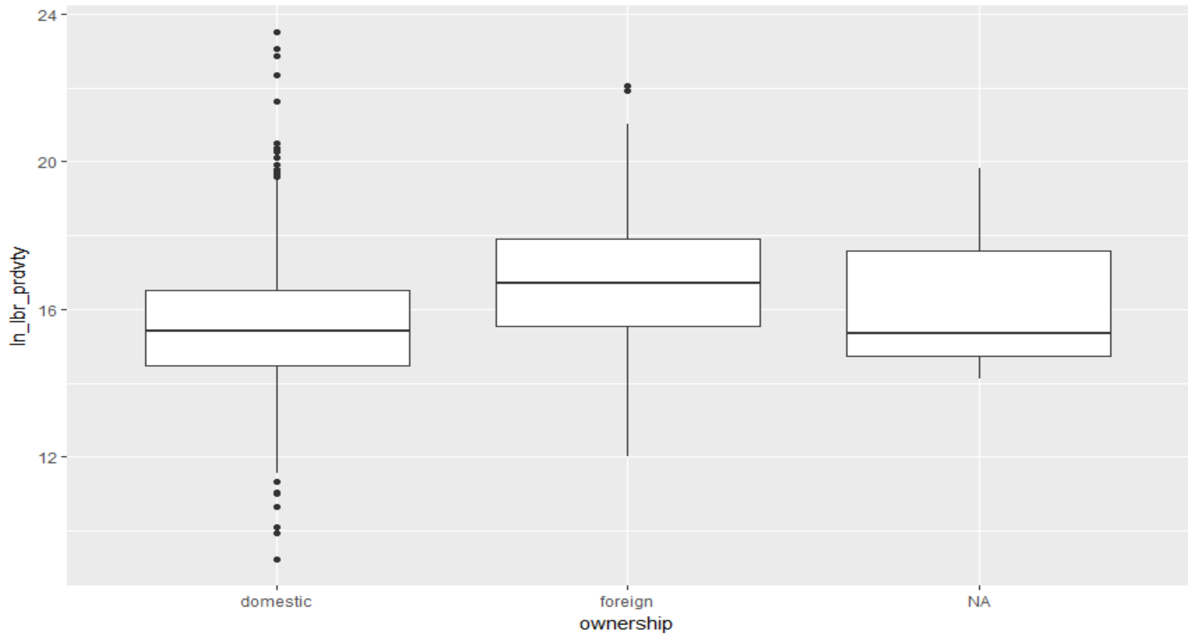


Figure 2. 1: Proportional difference between foreign and domestically-owned firms on labour productivity dimension

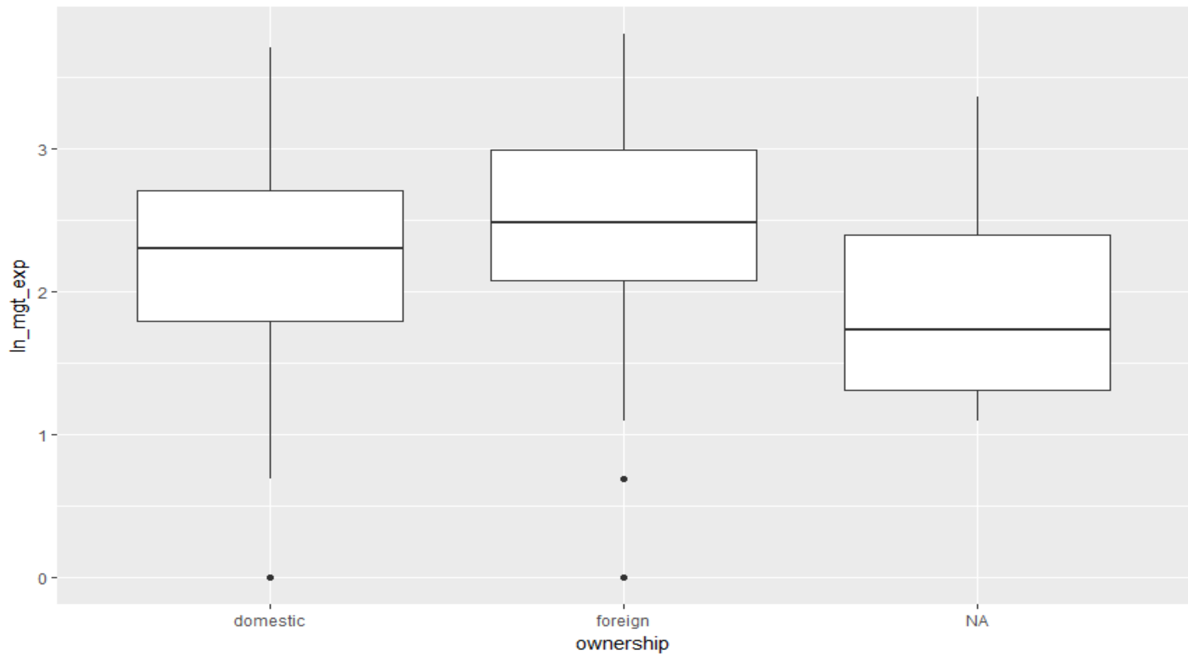


Figure 2. 2: Proportional difference between foreign-owned and domestically-owned firms on management experience dimension

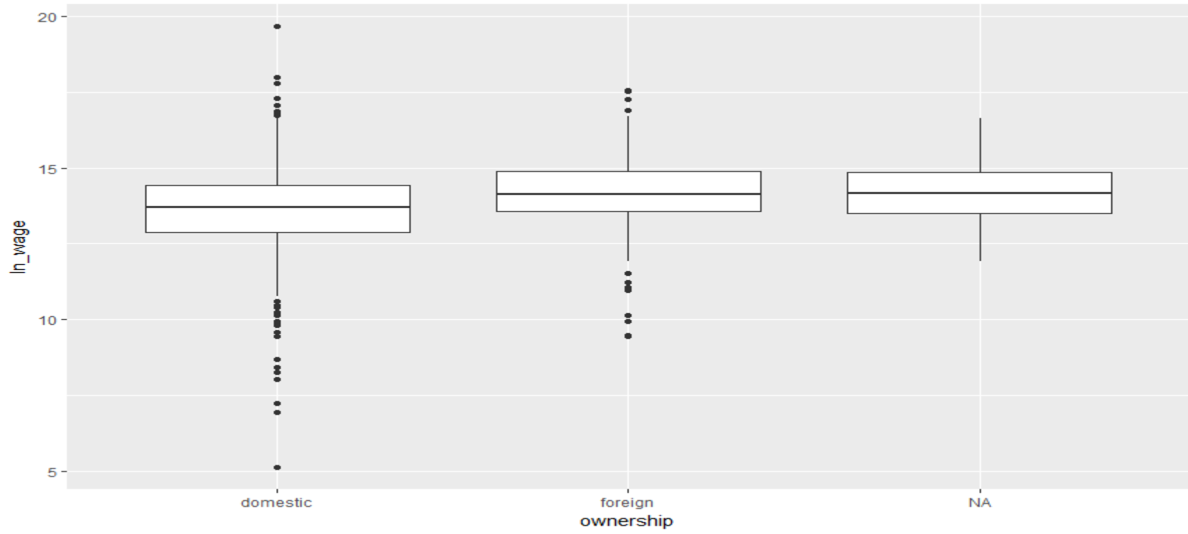


Figure 2. 3: Proportional difference between foreign-owned and domestically-owned firms on average wage dimension

The results from both descriptive statistics and regressions provide insights into some of the key characteristics of firms in Uganda. They also provide preliminary evidence regarding the belief that foreign-owned firms are likely to differ from domestically-owned firms. This insight is evident even when foreign-owned firms are indicated to be proportionately fewer than their domestically-owned counterparts. In Figures 2.4 and 2.5, it is illustrated how proportionately few foreign-owned firms exist relative to domestically-owned firms; and how these foreign-owned firms break down into the various categories.

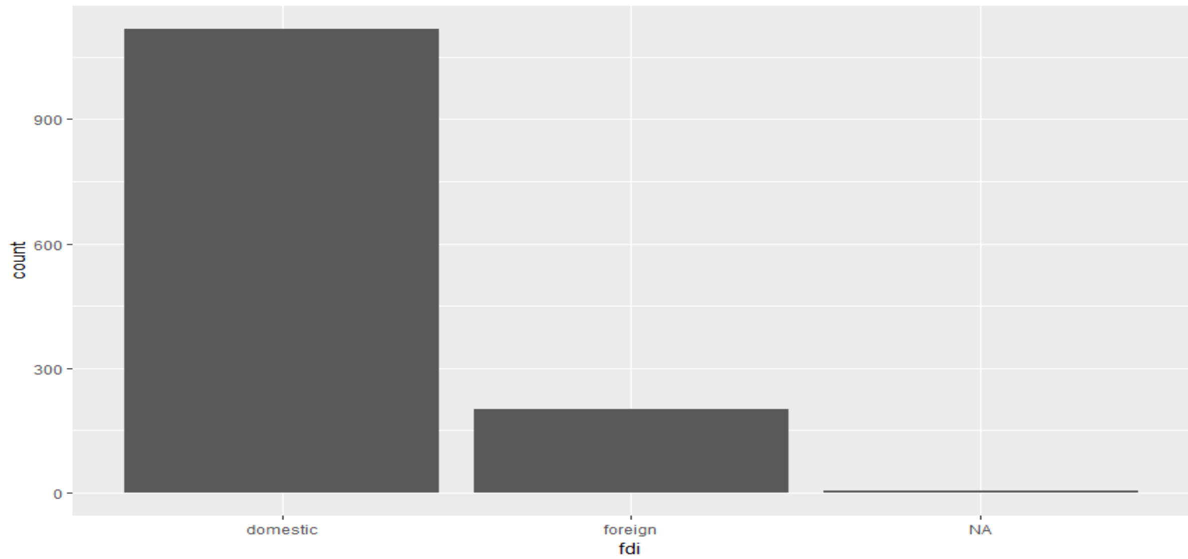


Figure 2. 4: Proportion of foreign-owned firms based on overall data

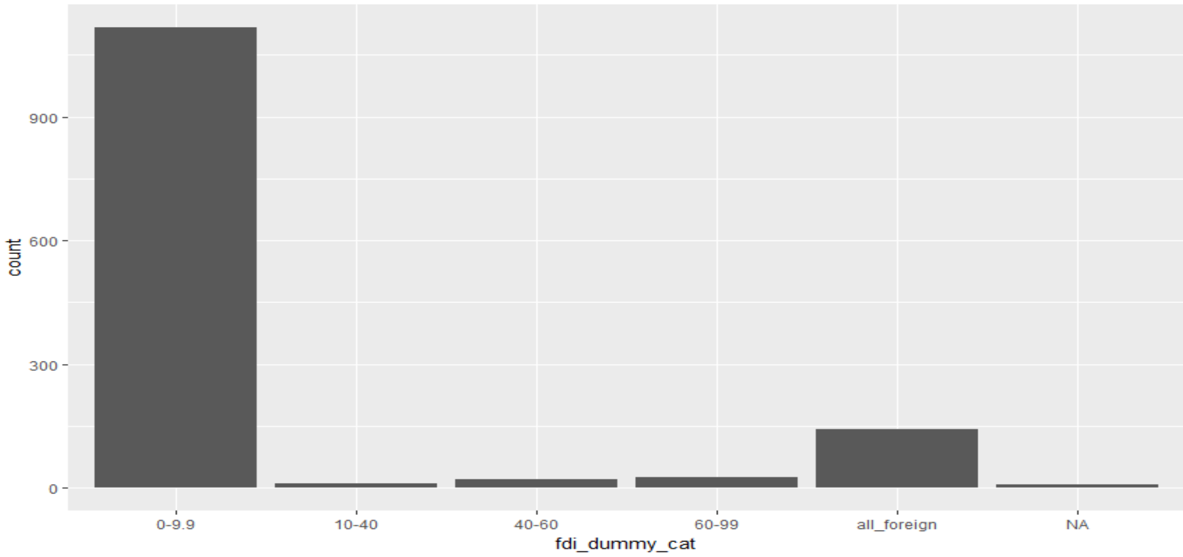


Figure 2. 5: How foreign-owned firms break down by percentage share category

Figure 2.4 highlights the fact that foreign-owned firms are about 200 and represent about 15% of the total number of firms in the survey. In figure 2.5, the breakdown by category of these foreign-owned firms is illustrated. We notice that the tallest bar corresponds to 0-9.9 category and going by the OECD definition of foreign ownership, which guides this thesis, this category is constituted by domestically-owned firms. The rest of the bars except the last on the right, correspond to foreign-owned firms and it is evident that the largest percentage of foreign-owned firms is composed of those whose ownership is 100% foreign, suggesting that ownership probably emerged from either complete takeovers or totally new plant being established. Categories 40-60% and 60-99% are proportionately similar with very few firms falling into the 10-40% category. The large percentage of completely foreign-owned firms lends more credence to the analysis.

The first stage of analysis provided preliminary insights into the salient characteristics of foreign and domestic firms in Uganda. These results further highlighted the preliminary fact that foreign-owned firms and domestically-owned firms are likely to differ on a number of dimensions. However, it was still unclear and hence challenging to conclude whether the observed differences between foreign-owned firms and domestically-owned firms were due to comparisons between different sub-groups (for example, between large firms where foreign-owned firms are more prevalent and small firms where they are not) or whether foreign-owned firms are different systematically across a number of dimensions. It was this challenge that led to the second stage of analysis, the hierarchical cluster analysis of the firms.

2.6.2.3 Standard Hierarchical Clustering

This is the second stage of analysis in this chapter. Methodologically, I began with an assessment of whether the data I intended to cluster could be clustered. This action is what cluster analysts refer to as Cluster Tendency Assessment (CTA). CTA is essential because clustering algorithms will always yield clusters, even when the data does not indicate so. Data may simply contain objects that are randomly distributed and hence unfit to be clustered. It makes less sense to subject a data set without naturally inherent clusters to any kind of partitioning (Banerjee & Davé, 2004). Therefore, CTA is useful because it indicates to the analyst whether the available data contains inherent clusters and is therefore amenable to cluster analysis. Two methods of CTA are popular in the literature, namely statistical methods and visual methods. The most common statistical method is the Hopkins statistic (H), whereas the most common visual method is the Visual Assessment of cluster Tendency (VAT). VAT is challenging when there are many items to cluster, such as in this study. I therefore utilised the Hopkins (H) statistic to perform the CTA.

The Hopkins Statistic (H)

The H statistic measures the likelihood that a given data set is generated by a uniform data distribution. According to Charrad *et al.* (2014), the H statistic algorithm consists of five steps as briefly described below:

1. *Step 1*: Sample uniformly n (p_1,…….p_n) points from the data set to be used, for example K.
2. *Step 2*: For each point $p_i \in K$, find its nearest neighbour p_j; and compute the distance between p_i and p_j. This can be denoted as $x_i = dist\ (p_i, p_j)$.
3. *Step 3*: Generate a simulated data set (random K) drawn from a random uniform distribution with n points (q_1, ………, q_n) and the same variation as the original real data set K.
4. *Step 4*: For each point q_i, find its nearest neighbour q_j in K, and then compute the distance between q_i and q_j and denote it as y_i = dist (q_i, q_j)
5. *Step 5*: Calculate the H statistic as the mean nearest neighbour distance in the random data set relative to the sum of the mean nearest neighbour distances in the real and across the simulated data set.

The H statistic is specified as:

$$H = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i} \tag{2.8}$$

Interpretation

If H is close to 0.5, it implies that $\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} y_i$ are close to each other and hence data set K is uniformly distributed. The null hypothesis is; the dataset K is uniformly distributed and therefore does not contain meaningful clusters. The alternative hypothesis is; the data set K is not uniformly distributed and hence contains clusters that are meaningful. If H is close to zero, then we reject the null hypothesis and conclude that data set K is indeed clusterable. In this chapter therefore, I perform the CTA and proceed to cluster analysis.

I clustered the firms first, based on selected variables. Then I clustered the variables themselves to help generate groups of variables with similar information and gain insights into how far the correlation between variables might affect clustering results of firms. Scholars such as Naughtin and Rankin (2016) have successfully employed similar analytical procedures in segmenting and analysing characteristics of South African exporting firms. When clustering firms, I used the hierarchical agglomerative method. Due to its many advantages, I preferred this clustering algorithm over others, such as the K-means, self-organising maps, and expectation maximizing clustering algorithms. One such advantage is that, in cases where the data sets are smaller such as the one used in this chapter, hierarchical algorithms show more quality and accuracy when compared to other algorithms such as the K-means and self-organising maps algorithm (Manpreet kaur & Usvir Kaur, 2013). Additionally, hierarchical algorithms are associated with embedded flexibility regarding granularity, with ease of handling any form of similarity or distance, and they are more versatile (Abbas, 2008).

I achieved similarity between firms by first estimating the distances between them. I estimated these distances using Gower's general similarity coefficient (equation 2.4) because it takes care of mixed data and has varying components of variables [12]. Since the data set I had used for analysis consists of variables that are mixed, this coefficient was thought to be suitable. This suitability is derived from its ability to handle various types of data such as ordinal, nominal and binary variables, even when they appear in the same data set.

[12] Gower's general similarity coefficient is an option in the daisy function of the R statistical software package, the computer tool used in this chapter's analysis.

After estimating the distances between firms, the firms were clustered via a linkage technique called Ward's linkage method. This linkage method is also called Ward's minimum variance method. It minimises the sum of squares to form clusters, which is the reason why it is also called the incremental sum of squares method. Ward's linkage method was selected because of recommendations by earlier studies that were reviewed, but also because it is applicable to weighted clustering, unlike other methods such as average linkage, complete linkage and single linkage. Hands and Everitt (1987) conducted a comparative study of linkage methods and concluded that Ward's linkage method performed better than other linkage methods applicable to hierarchical clustering. Blashfield (1976) also compared four types of hierarchical clustering methods (single linkage, complete linkage, average linkage and Ward's method) for accuracy in recovery of original population clusters. His results indicated Ward's method to be superior to other methods. A more recent comparative study of hierarchical linkage methods is by Ferreira and Hitchcock (2009), whose results also rank Ward's method ahead of other methods. The major justification, however, for use of Ward's linkage method in this study was the need to utilise its weighted clustering abilities, since the survey data had weights.

Other scholars have, however, presented the strengths of other linkage methods. Šulc and Řezanková (2014) compare three linkage methods for hierarchical clustering of categorical data, and recommend use of either the complete linkage or average linkage method. These methods provide good differentiation of clusters, making it easy to cut a dendrogram at a given point, among other merits. Moreover, average linkage embodies a natural compromise between single linkage and complete linkage, as it is sensitive to the shape and size of clusters (Yim & Ramdeen, 2015). Overall, hierarchical methods form the backbone of cluster analysis, are widely available in software packages and are easy to use, leaving the analyst with only a decision to make regarding proximity measure and number of clusters (Everitt *et al.*, 2011: 110).

After linking the firms and consequently generating a large number of firm clusters, the number of optimal clusters to use in further analyses had to be decided. Determination of the optimal number of clusters has been associated with subjectivity and based on expert judgement of the analyst in most cases. However, several methods have been developed to guide in determining the optimal number of clusters. The use of a dendrogram is one of the methods. This is a graphical illustration of how clusters are merged, allowing the analyst to identify the appropriate number of clusters. This provides a mathematical and pictorial representation of the complete clustering procedure, with nodes denoting the clusters and the length of the stems representing the distances at which clusters are joined (Everitt *et al.*, 2011). Each node has two edges (binary trees) emanating from it; and the objects in question form the labels at the terminal nodes of the dendrogram. To determine the optimal number of clusters, this tree-like diagram is cut at a point where the distances are relatively large. This method, however, is susceptible to great challenges as the number of observations becomes larger. Visual establishment of the point at which to cut the tree is an enormous task.

The agglomeration schedule is another method for selecting the number of clusters. The schedule shows the step-by-step process of the clustering. It is easily used and interpreted when the distances are plotted against the number of clusters, yielding a scree-plot graphical illustration. On the scree-plot, a distinct break-point, usually referred to as the elbow, will indicate the point at which joining further different clusters can only occur at greater distances. By implication, the number of clusters just before the elbow provides the solution to the optimal number of clusters to be considered by the analyst (Mooi & Marko, 2011). However, this method can provide ambiguous results in some instances. Both the agglomeration schedule and the dendrogram methods are thus associated with challenges.

To surmount some of the challenges associated with the aforementioned methods, several precise methods have been proposed in the form of indices. These indices also guide in determining the optimal number of clusters during the analysis. The indices also have their own strengths and weaknesses. Several scholars have examined these, with each scholar recommending particular indices with related justification. For instance, Milligan and Cooper (1985) compared 30 indices of determining the optimal number of clusters in hierarchical clustering algorithms. Their results ranked the Calinski and Harabasz's (CH) index as the best, followed by Duda and Hart's (DH) index, with the C-index in third place. In a simulation study based on eight-dimensional outcome variables taken from a real case study of schizophrenic patients, Islam *et al.* (2015) also compared several indices. They found that the DH index, Hartigan (H) index and Gap/pc index are the best performing indices. These two studies appear to provide contradictory results. However, Milligan and Cooper's findings have been supported by scholars like Yan (2005), Everitt *et al.* (2011) and Řezanková (2014). Because of the varying recommendations in scholarly literature, some analysts compute all the indices and choose the number of clusters produced by majority indices. This approach is called the majority rule decision method. In this chapter, both the CH and majority rule methods were used. Majority rule method helps to robustly support the CH index results[13].

The CH index is based on a variance ratio criterion proposed by Caliñski and Harabasz (1974) specified as below:

$$\mathrm{CH_k} = \frac{\mathrm{ss_B}/(\mathrm{k}-1)}{\mathrm{ss_W}/(\mathrm{n}-\mathrm{k})} \tag{2.9}$$

where *k* denotes the number of clusters, *n* denotes the number of objects (firms in this case), $\mathrm{ss_B}$ measures the overall variation between segments, and $\mathrm{ss_W}$ measures overall variation within segment with regard to all variables used in the clustering. The optimal number of clusters is the value of *k* that maximises $\mathrm{CH_k}$[14].

Once the optimal number of clusters had been determined, analysis of the characteristics of each cluster was undertaken and results presented and discussed. Each cluster's characteristics were then analysed using descriptive statistics. Comparisons between clusters were undertaken, differences of cluster scores on each variable were established, and their significance was statistically tested using students' T-tests. Regressions were also run on the variables, using clusters as factors to understand how these clusters are likely to be associated with variations in these variables. This regression result indicated which clusters were more likely than others to be associated significantly with certain performance indicators.

After systematically applying the above clustering methodology, the results that follow were generated.

2.5.2.2.2 Clusters of Firms

The analysis starts with a statistical test of the clustering tendency on the data sets. In Table 2.8, the Hopkins statistics are shown in the fourth column. In all the three sub data sets, the statistic is about 0.2. I hence rejected the null hypothesis that the data is uniformly distributed and therefore does not contain meaningful clusters.

[13] The R package clusterSim also contains some of the more recently introduced criteria including silhouette index and GAP- statistics.

[14] In the R software, the NbClust yields the optimal number of clusters as determined using the CH index.

Table 2. 8: Cluster solutions and clustering tendency for the sub data sets

Data set	Optimal Clusters	Value Index	Hopkin's Index
NData2006	2.0	466.01	0.2618
NData2013	3.0	461.58	0.2713
CNData	3.0	809.16	0.2382

Source: Author's own results based on World Bank Enterprise Survey Data, 2006/13 panel

Turning to a standard hierarchical clustering of firms using agglomerative procedures, the distance between firms was measured using Gower's distance measures. Thereafter, firms were linked using Ward's 'Ward.D2' method. Ward's linkage method has the ability to integrate weights in the clustering process. A weighted clustering for the three sub data sets (NData2006, NData2013 and CNData) was therefore performed.

As indicated earlier on, the optimal number of clusters was selected, guided by visual illustrations like the dendrogram, Silhouette methods, majority rule of indices, and the Caliñski and Harabasz's index. In Figure 2.6, the dendrograms demonstrate the arrangement of clusters produced by the corresponding analyses. On the x-axis are the firms that are clustered while on the y-axis are the distances at which the clusters are formed. A horizontal line drawn from one of the distances on the y-axis to the left will cross some vertical lines on the dendrograms. Each vertical line crossed corresponds to a cluster. For instance, with reference to the first dendrogram, a line drawn from 2.0 on the y-axis to the left, will cross two vertical lines yielding a two-cluster solution. A visual inspection of dendrograms for the three sub data sets suggest between 2 to 5 clusters as shown by the coloured boxes on each diagram.

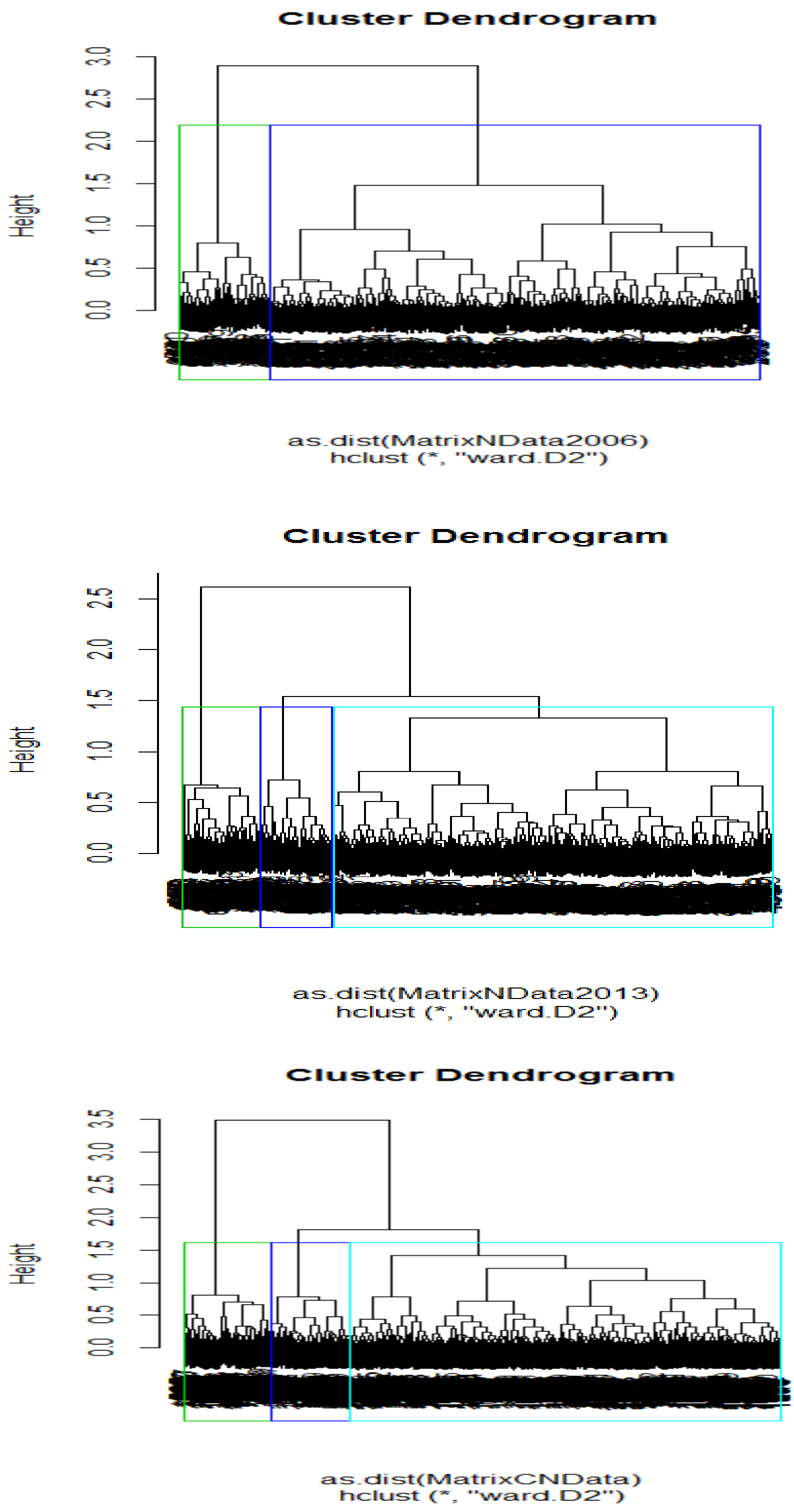


Figure 2. 6: Cluster dendrograms for NData2006, NData2013 and CNData

In Figure 2.7, we notice the silhouette illustration for 2006 firms. The average silhouette width across all firms shows the overall quality of the clustering result and this is indicated on the y axis. On the X axis is the number of clusters. A larger averaged silhouette width shows a superior overall quality of the clustering result and from the illustration, the largest is just slightly below 0.25, which corresponds to 2 clusters. However, an average silhouette width of 0.25 shows that the structure is weak and calls for additional methods of analysis to be explored. This is why I utilised the indices i.e. the CH index and the majority rule method.

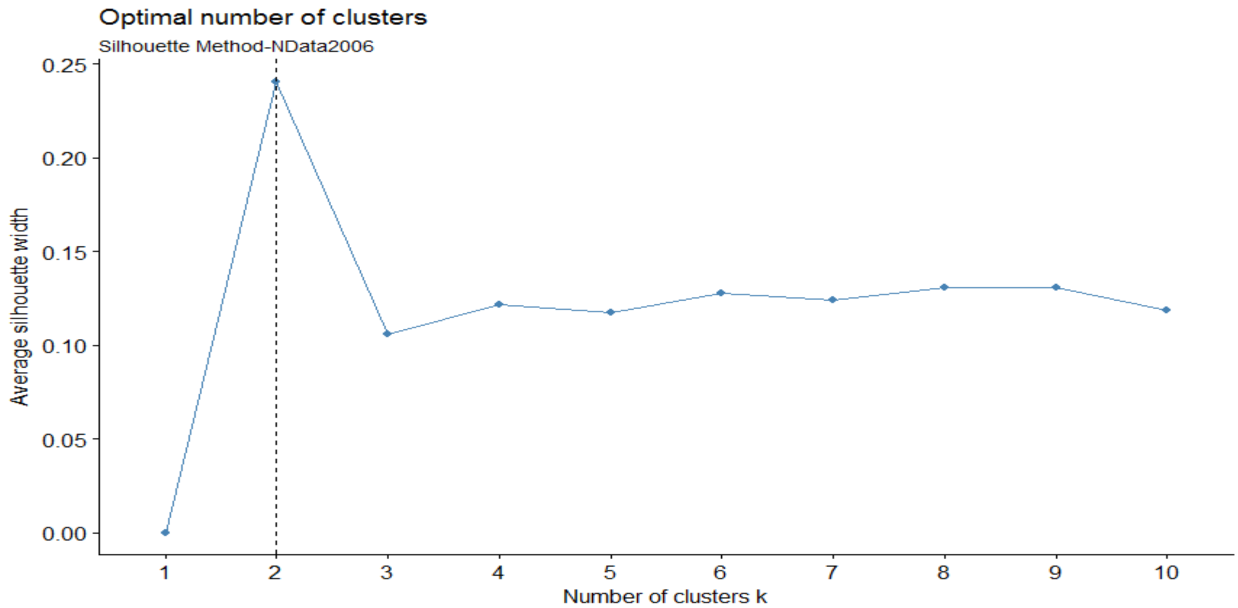


Figure 2. 7: Silhouette method guide to optimal cluster selection

In Figure 2.8, the silhouette illustration for overall data shows that the largest average silhouette width is about 2.5. which calls for other methods to be explored. However, silhouette method suggests 2 clusters in the overall data set. The elbow method suggests 4 clusters for the 2013 data.

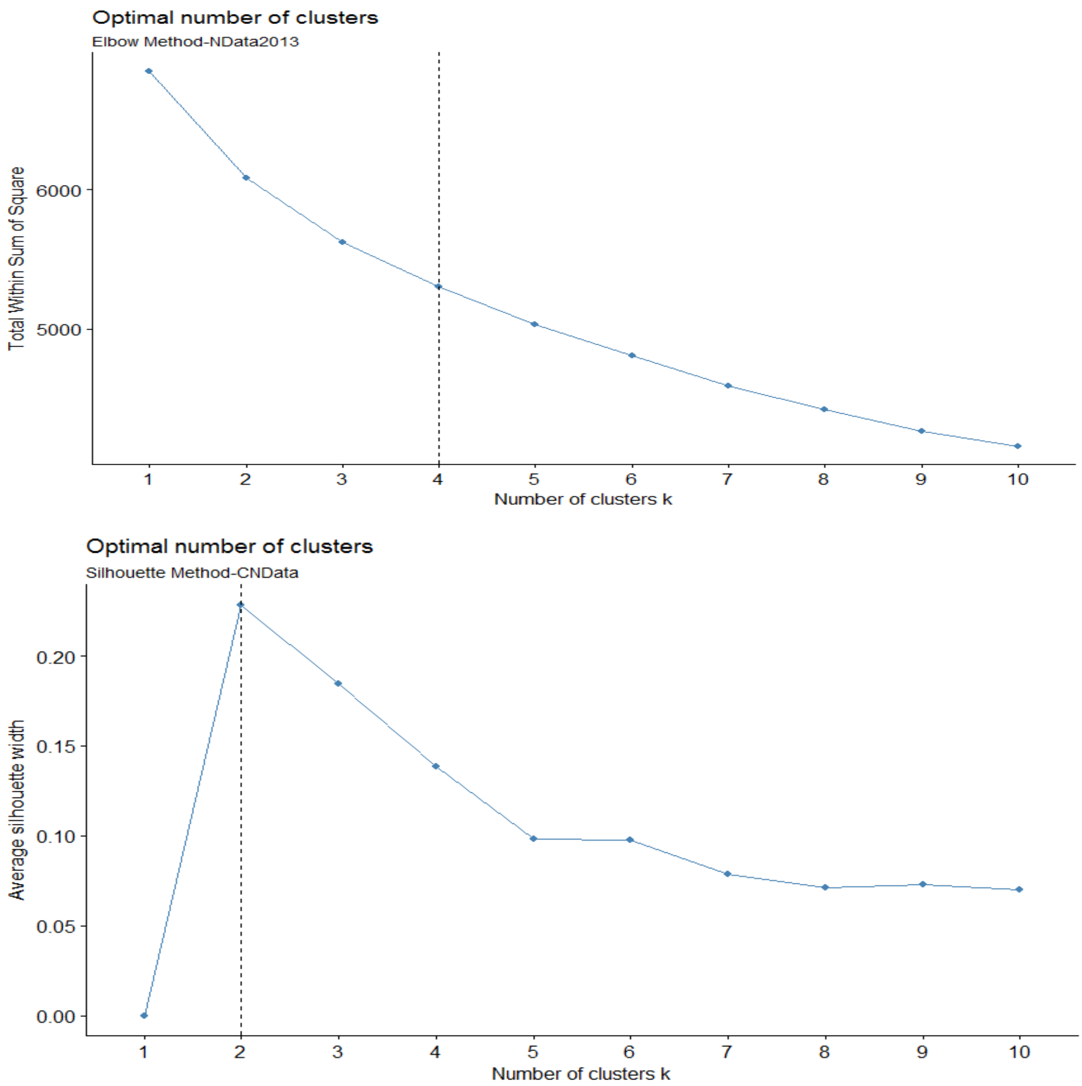


Figure 2. 8: Silhouette and elbow methods for respectively 2013 and combined data

However:

> *"...determining the optimal number of clusters in a data set is a fundamental issue in partitioning clustering, such as K-means clustering, which requires the user to specify the number of clusters k to be generated. Unfortunately, there is no definitive answer to this question. The optimal number of clusters is somehow subjective and depends on the method used for measuring similarities and the parameters used for partitioning. A simple and popular solution consists of inspecting the dendrogram produced using hierarchical clustering to see if it suggests a particular number of clusters. Unfortunately, this approach is also not without some subjectivity"* (Kassambara, 2017: 128)

In order to reduce such subjectivity, I turned to the majority index rule method and the Caliñski and Harabasz index. In Table 2.8, the optimal number of clusters for the three sub-data sets together with the associated value indexes are presented. The distances on the dendrograms were cut and marked based on results in Table 2.8.

Also, in Table 2.8, it is noticeable that the majority rule decision and the CH index yield 2, 3 and 3 clusters respectively for the 2006, 2013 and combined data. The combined data set yielded a three-cluster solution – as was the case for 2013 – but with a high value index and lowest Hopkins index (highly clusterable structure). The results were obtained using the NbClust R package[15]. Using this package, I was able to run over 20 indices and to detect what majority of scholars identify as the optimal number of clusters. Some of these indexes include but are not limited to: Krzanowski and Lai (1988), Calinski and Harabasz (1974), Duda and Hart (1973), Ratkowsky and Lance 1978, Frey and Van Groenewoud 1972, Halkidi and Vazirgiannis 2001, Halkidi et al. 2000, Tibshirani et al. 2001, Rousseeuw 1987, Milligan and Cooper 1985 and Hartigan 1975. It was additionally noticed that both the CH and majority indexes provided cluster numbers that were within those suggested by the dendrograms, the elbow, and silhouette methods.

[15] Developed and published by Charrad, Ghazzali, Boiteau and Niknafs (2015)

2.6.2.4 Characteristic Features of Firms' Clusters

Clusters of the combined data set: *CNData*

Table 2.9 shows the weighted mean and median real values of numeric variables for the three clusters of the overall data set.

Table 2. 9: Descriptive Statistics, Weighted Mean and Median (real) values CNData

Variable	Cluster 1, N = 193	Cluster 2, N = 178	Cluster 3, N = 954
log employment	3.06	2.28	2.32
log labour productivity	17.15	15.93	15.37
fdi stake	92.74	0.13	0.06
log capital intensity	14.95	13.92	13.97
log material/worker	15.77	14.86	14.17
log management exp	2.34	2.00	2.15
log wage	14.07	13.40	13.45
log worker education	2.46	1.10	2.53
log absolute exports	21.06	18.59	18.01
actual employment	41.77	15.66	22.21
Median			
log employment	3.00	2.20	2.08
log labour productivity	16.67	15.62	15.25
fdi stake	100.0	0.00	0.00
log capital intensity	15.32	14.17	13.85
log material/worker	15.88	14.71	14.11
log management exp	2.40	1.95	2.30
log wage	13.96	13.74	13.54
log worker education	2.64	1.10	2.56
log absolute exports	21.74	17.16	17.96
actual employment	20.00	9.00	8.00

Source: Author's own computations based on World Bank Enterprise Survey, 2006/13 panel

Table 2.10 shows the results of the three clusters for qualitative variables in the overall data set.

Table 2. 10: Descriptive statistics of qualitative variables for CNData clusters

Variable	Cluster1, N = 193	Cluster 2, N = 178	Cluster 3, N = 954
sector type			
manufacturing	26.2%	14.4%	27.2%
Retail	18.8%	15.2%	35.8%
Others	55.0%	70.4%	37.0%
firm size (№ of workers)			
large>=100	9.6%	1.1%	2.3%
small>=5&<=19	51.7%	76.2%	83.1%
medium>=20&<=99	38.7%	22.8%	14.6%
firm age			
<10	51.4%	55.4%	51.9%
>=10&<=50	46.4%	44.0%	47.6%
>50&<90	2.2%	0.6%	0.5%
R&D Expenditure			
No	60.3%	64.9%	72.1%
Yes	39.7%	35.1%	27.9%
formal employee training			
No	58.7%	58.9%	67.2%
Yes	41.3%	41.1%	32.8%

Source: Author's own computations based on World Bank Enterprise Survey, 2006/13 panel

According to the results in Tables 2.9 and 2.10, these clusters can be described as follows:

Cluster 1: Slightly large/medium, highly productive and capital-intensive firms (14.6% of the sample)

It is noticeable that cluster 1 has a 92.74% fdi stake on average compared to cluster 2 and 3 with less than 1%. This high fdi stake is suggestive of a cluster where foreign-owned firms are more likely to be found. Results in Table 2.8 indicate that mean and median scores of this cluster on all variables are above those of both cluster 2 and 3, the only exceptions being; median wage, and worker_educ. Cluster 1 firms employ on average 41.77 workers ('log employment' is also is 3.06) and have a median value of actual employment of 20 workers. This result suggests that firms in this cluster are more likely to be medium- but also possibly large-sized firms. In Table 2.9, it is noticeable that 9.6% of firms in this cluster are larger in size and 38.7% are medium-sized firms amounting to a combined percentage of 48.3%. Although 51.7% are under the small-size category, the 48.3% outweighs a similar combined percentage in both clusters 2 (23.9%) and 3 (16.9%). In Figure 2.10, it is noticeable that, besides cluster 1, large firms are clustered too in cluster 3; and that all clusters have presence in the medium- and small-size categories. However, the combined percentage above separates cluster 1 from clusters 2 and 3.

Noticeable too in Table 2.10 is that firms in this cluster are more likely to be in other sectors than manufacturing and retail. They post a higher percentage (55%) in other sector categories but only 26% in manufacturing and 18.8% in retail. The lower percentage in manufacturing for a cluster where foreign-owned firms are likely to be grouped contradicts findings from earlier studies on Uganda. One such a case is Obwona (1998: 18) in which over 70% of foreign direct investment is found to be destined for manufacturing. Indeed cluster 2 firms post a 27.2% likely presence in manufacturing yet it's a cluster likely to have more domestically-owned firms.

The percentage of cluster 1 firms that fall in larger category (>=100 workers) is higher compared to firms in cluster 2 (1.1%) and cluster 3 (2.3%). This result is, to some extent, consistent with the mean values (Table 2.11) of actual employment of cluster 1(41.77) that are higher than those of cluster 2 (15.66) and cluster 3 (22.21). Table 2.9 also shows that cluster 1 is likely to have firms with higher levels of productivity. This likelihood is indicated by labour productivity of 17.15 and material used per worker at 15.77. Indeed, in Table A3 in the appendices, it is noticeable that firms in cluster 1 have comparatively higher values of real output.

Elsewhere, although firms do not generally prioritise expenditure on R&D, firms in this cluster post a higher percentage at 39.7% compared to 35.1% and 27.9% for cluster 2 and 3 respectively. This result suggests that foreign-owned firms are more likely to spend on R&D than their locally owned counterparts. This result is in agreement with the literature; and provides justification for the argument that foreign-owned firms are better at new product development than domestic firms. This is probably due to the characteristic of export orientation that requires frequent improvements to keep up with international markets' standards. Indeed, cluster 1 posts higher mean and median values (Table 2.9) for absolute exports (21.06 & 21.74 respectively) compared to cluster 2 (18.59 &17.16 respectively) and cluster 3 (18.01&17.96 respectively). In Table 2B in appendices we see clusters' actual exports.

Similarly, cluster 1 has higher mean/median values for capital intensity (14.95/15.32) compared to cluster 2 (13.92/14.17) and cluster 3 (13.97/13.85). Cluster 1 also differs from specifically cluster 2 in terms of; management experience, worker_educ, and exports. These differences are significant as indicated by the respective p-values in Table 2A in the appendices. This result is in agreement with the popular argument in literature that foreign-owned firms are likely to be export oriented (Erdal & Göçer, 2015b; Kimura & Kiyota, 2006), a feature that links them to use of more skilled (more highly educated on average) workers (Matsuyama, 2007 as quoted by (Brambilla, Lederman & Porto, 2012: 3406), and experienced management compared to domestically owned firms. These results are specifically, in agreement with Moss and Ramachandran (2004: 3) in their study on foreign direct investment in East Africa. Results on worker_educ, however, contradict some earlier studies on foreign direct investment in Uganda, especially regarding managerial workers who were found to be wanting in terms of education (Barthel *et al.*, 2008; Obwona, 1998: 21).

The features associated with cluster 1 in the combined data set's results, to a great extent, mirror those associated with clusters 1 and 2 in 2006 and 2013 respectively. Tables 2H and 2I in the appendices contain clustering results based on 2006 and 2013 sub data sets. The two clusters, results show, have fdi stake averagely at 93.42% and 94.23% respectively. Moreover, the median of fdi stake of 100% suggests that these clusters are more likely to include some firms that are completely foreign. Although with slight variations, these clusters in their respective sub data sets (and hence tables) have higher mean and median values on most variables compared to those cluster where foreign-owned firms are less likely to be grouped. The slight differences visible in Tables 2H and 2I are attributed to differences in sample sizes.

Cluster 2 and Cluster 3 CNData: *Largely small- and medium-sized, moderately skill- and Capital-intensive firms*

These clusters form respectively 13.4% and 72% of the overall data set. These clusters further have mean fdi_stake of less than 1% each according to Table 2.9. This low fdi stake is suggestive of clusters in which foreign-owned firms are less likely to be found, but domestic firms. In Table 2.9, cluster 2 firms employ on average 15.7 workers (log employment also is 2.28) and have a median value of actual employment of 9 workers. This result is suggestive of a cluster in which largely small firms are grouped. In Table 2.10, this cluster has 76.2% under the small size category and this result resonates well with the histogram caption in Figure 2.12. The 1.1% result on large firm size category corroborates what we see in figure 2.12 for this cluster. Comparatively cluster 3 has a 22.21 value on actual employment (log employment is 2.32) but a lower median value of 8. This result is suggestive of a cluster likely to have smaller sized but also medium sized firms. In Table 2.10, this cluster has 83.1% under small size category and 14.6% for medium sized firms. These two clusters have relatively similar firms regarding size. Most of their firms are mainly small and medium sized although cluster 3 tends to have more comparative presence in the large size category. Figure 2.9 contains a histogram for the three clusters and all firm size categories. The illustration confirms the relative cluster comparisons alluded to above. Cluster 1 with predominantly foreign-owned firms has firms in all the three categories, but mainly in the large and medium. The largest number of firms in cluster 3 are small in size.

A notable characteristic that separates these cluster 2 and cluster 3 more evidently is that of worker_educ. On average, whereas the log of worker education stands at 1.1 for cluster 2, for cluster 3 it is at 2.53. Table A2 in the appendix also indicates that between cluster 2 and cluster 3, the p-value on education is the most different from the rest. This observation tends to suggest that firms in cluster 3 are more likely to be skill intensive than those in cluster 2. However, intriguingly, cluster 2 seems more exporting (18.59) in absolute terms than cluster 3 (18.01), although this difference is not statistically significant as per the p-value in Table A2.

Cluster 3 also differs from cluster 2 in terms of firms' management experience as par Table 2.9. Additionally, Table 2.10 shows this cluster having higher values in favour of R&D expenditure and likelihood of providing formal training to firm employees. 72% of firms in cluster 3 are likely to spend on R&D compared to 64.9% firms in cluster 2. Also 67.2% firms are more likely to offer formal training to their workers compared to 58.9% in cluster 2. Firms in cluster 2 are mainly in retail and other sector categories with less presence in manufacturing (14.4%) compared to both cluster 1 (26.2%) and cluster 2 (27.2%). R&D is mainly a priority for firms that are anchored in the manufacturing sector where product development is critical so as to maintain market share as opposed to firms in retail whose main activity is distribution and sale of what has been produced. These results imply that most domestic firms in Uganda are not research intensive and hence far below the technology frontier. They are usually either small or medium enterprises as confirmed in figure 2.9. This automatically feeds into their performance abilities when compared to their foreign-owned counterparts.

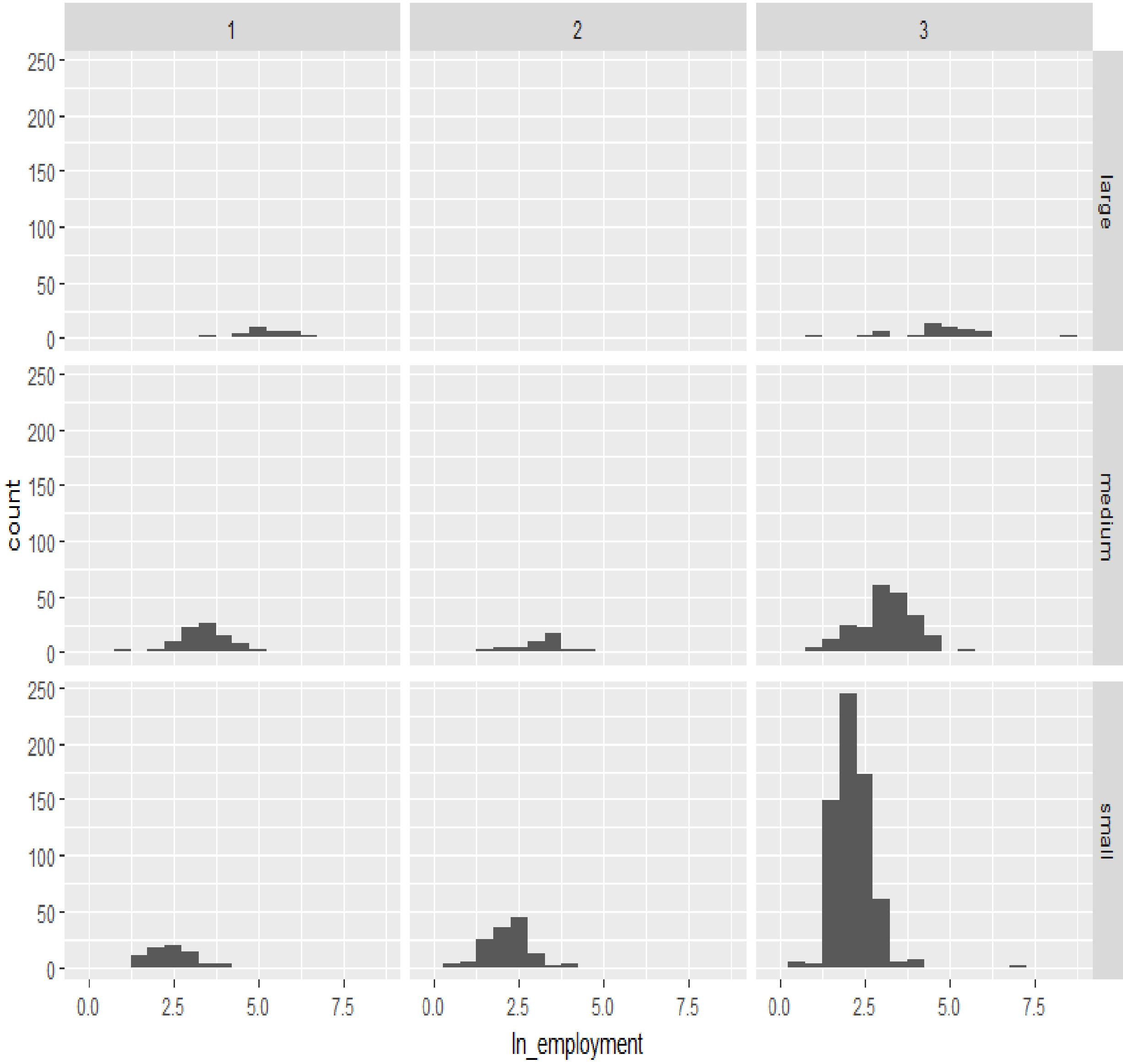


Figure 2. 9: Histograms of firm size by cluster based on overall data

In this chapter, cluster analysis makes use of Euclidean distance measures. Whereas use of this measure is admittedly popular and acceptable, it is susceptible to one natural problem that may be associated with variables used and hence distort results. This problem is "variable collinearity". While the problem of multicollinearity is easy to see in regression analyses by looking at unreasonable beta coefficients in the results output, it is challenging in cluster analysis where betas and dependent variables don't exist. When some variables used in clustering are highly collinear, they don't add much new information to the description of the clusters. Collinearity causes information reduction and leads to cluster descriptions that aren't as rich as they could have been. In practice, good clusters are usually those with a mixture of high, low, and moderate values on different variables (Sambandam, 2003: 19). But existence of high collinearity or correlation among variables (or some) in the data, makes mixed-value clusters extremely hard to achieve.

Several methodological solutions have been developed to mitigate the problems associated with multicollinearity of variables in cluster analysis. One method that probably is applicable to this thesis is that of adopting the Mahalanobis distance measure other than Euclidean distance measure. Mahalanobis distance measure is not affected by variable collinearity (Pacáková & Poláčková, 2013: 54). But calculation of such distances involving many variables can be intricate and iteratively time intense (Sambandam, 2003: 20). Another easy method involves eliminating some of the correlated variables from the data before analysis. But this elimination is premised on the condition that the retained variables are practically useful and have high actionability potential, which might be difficult to ensure (Sambandam, 2003: 22).

One other popular technique one can use to surmount the correlation problems above is reduction of the dimension of the data so as to generate "*principal variables*" that are uncorrelated. This is popularly performed via either Principal Component Analysis (PCA) or Multiple Correspondence Analysis (MCA) for numeric or non-numeric variables respectively[16]. PCA/MCA involves a transformation process that leads to a reduction of the dimension of the data to a lower one with principal variables that are uncorrelated (Pacáková & Poláčková, 2013: 54). This analytical transformation is done while at the same time maximizing the magnitude of information recovered from the initial space. But PCA/MCA and other methods like diametrical clustering are either numerically or non-numerically leaning, and may not be suitable for data that is mixed (Chavent, Kuentz, Liquet & Saracco, 2011).

A relatively novel technique that is neither numerically leaning nor otherwise is the ClustOfVar approach. In order to generate cluster descriptions that are desirable or to at least provide a validation test to the clustering results, we, in this chapter, employ the ClustOfVar technique (other than PCA/MCA) in generating uncorrelated principal variables. These principal variables, which are also popularly called synthetic variables, are later used to re-cluster the

[16] PCA/MCA is a data analysis tool that is used to reduce the dimensionality (number of variables) of a large number of interrelated variables while retaining as much of the information (variation) as possible. PCA is mainly for numeric variables while MCA is mainly for non-numeric variables.

firms to see if similar cluster solutions are generated. Choice of this technique is informed by the challenges of PCA/MCA and other techniques aforementioned. ClustOfVar is a relatively novel technique with ability to surmount most of the challenges mentioned earlier. For instance, it works with a mixture of numerical and non-numerical variables and can as well work for a set exclusively containing numerical or non-numerical variables. Moreover, it allows for missing data, which are replaced by means of numerical variables and by zeroes for qualitative variables in the indicator matrix (Chavent *et al.*, 2011). Survey data usually has missing values. I therefore performed another related analysis involving clustering variables for the sole purpose of validating earlier results obtained from hierarchical clustering of firms. The associated procedures and results are presented next;

2.6.3 Clustering of the Variables-

Clustering of variables is a way to arrange variables into homogeneous clusters (groups of variables) that are strongly related to each other and thus bring similar information (Chavent *et al.*, 2011). If correlation amongst variables has insignificant effect on clustering results, then re-clustering using the clustered/variable groupings (principal variables) should yield similar number of firm clusters.

In this chapter, the ClustOfVar technique utilises the hierarchical clustering algorithm, which is based on PCAMIX, a principal component method for a mixture of numerical and non-numerical variables. The aim of the clustering algorithm is to maximise a homogeneity criterion. Homogeneity of a cluster is measured by the extent to which variables constituting such a cluster are linked to a central quantitative synthetic variable. Chavent *et al.* (2011) propose measuring this link using the squared Pearson correlation for numerical variables and the correlation ratio for qualitative variables.

The hierarchical algorithm can be defined in the following steps;

Let $\{x_1, x_2, \ldots\ldots., x_{p1}\}$ be a set of p_1 quantitative variables and $\{z_1, z_2, \ldots\ldots., z_{p2}\}$ a set of p_2 qualitative variables. And let X and Z be respectively the corresponding quantitative and qualitative matrices of dimension n x p where n is the number of observational units. The steps of the algorithm are;

Step 1: Start with P = p1 + p2 partitions. This means that in the first step, each variable constitutes a cluster of its own.

Step 2: Merge the two clusters that have the smallest dissimilarity (maximum similarity) between them. Given clusters K_1 and K_2, this dissimilarity condition is given by

$$(K_1, K_2) = H(K_1) + H(K_2) – H(K_1 \cup K_2) \tag{2.9}$$

The homogeneity (H) of cluster C_k is defined by Chavent *et al.* (2011) as below;

$$H(C_k) = \sum_{x_j \in C_k} r^2_{x_j, y_\kappa} + \sum_{z_j \in C_k} \eta^2_{y_k|z_j} = \lambda_1^k \tag{2.10}$$

Where r^2 is the squared Pearson correlation and η^2 is the correlation ratio and y_k is the cluster's central synthetic quantitative variable, a variable most linked to all variables in the cluster. y_k represents the first principal component of the PCAMIX applied to the standardised variables of C_k. Therefore, the first term of (2.10) measures the link between quantitative variables in C_k and y_k while the second term measures the link between the qualitative variables in C_k and y_k. Cluster homogeneity is at a maximum when all the quantitative variables are correlated or anti-correlated to y_k and when the correlations ratios of all qualitative variables are equal to 1. λ_1^k is the first eigenvalue of PCAMIX applied to the k cluster C_k of a partition say P_k from the algorithm. The algorithm therefore tends to merge the two clusters that result into the smallest decrease in H.

Step 3: The process should be stopped once all the variables are merged into one single cluster.

Just as the optimal number of clusters was determined when clustering firms, a similar step is performed in variable clustering. The ClustOfVar R package, however, provides the option of bootstrapping to help both in determining optimal cluster numbers and evaluating the stability of the partitions of variables (Chavent *et al.*, 2011). As earlier mentioned, once these variable clusters are generated and the scores of each group obtained in form of what is referred to as synthetic variables, I cluster the firms again based on these now few uncorrelated principle variables and compare results. The groupings of these variables are also presented and briefly discussed. The results of this stage of analysis are presented next, beginning with test evidence of variable collinearity.

Table 2.11 shows correlation results, based on overall data set, for the variables used in the clustering of firms. In order, from 1 to 9, these abbreviated variables are; log employment, fdi_stake, log labour productivity, log wage, log management experience, log capital intensity, log material per worker, log worker education, and log absolute exports.

Table 2. 11: Correlation matrix for variables used in clustering firms in overall data set

Variables	1	2	3	4	5	6	7	8	9
1. empt	1.00								
2.fdi_stake	0.27	1.00							
3.prdvty	0.17	0.26	1.00						
4.wage	0.04	0.14	0.55	1.00					
5.mgt exp	0.19	0.10	0.04	0.02	1.00				
6.k-int	0.22	0.22	0.46	0.43	0.14	1.00			
7.mat/work	0.10	0.21	0.71	0.66	-0.06	0.42	1.00		
8.educ	0.17	0.12	0.04	0.05	0.09	0.12	0.01	1.00	
9.exports	-0.07	-0.04	0.01	-0.07	-0.14	-0.15	-0.22	-0.37	1.00

Source: Author's own illustration based on World Bank Enterprise Data 06/13 for Uganda

Although majority of the correlation coefficients are within the recommended (<0.5) level, some are either above or too near that threshold. Specifically, labour productivity is correlated with wage, capital intensity, and material per worker at respectively 0.55, 0.46, and 0.71 levels. The wage variable is highly correlated with capital intensity and material per worker at 0.43 and 0.66 respectively while the correlation coefficient between capital intensity and material per worker, 0.42, is too close to the threshold. Figure 2D in the appendices provides a visual illustration of these correlations. The plot in this figure mirrors largely the matrix. In Table 2C in the appendices, p-values of some of these correlations indicate statistical significance. This significance of variable collinearities, justifies the need for variable clustering to examine if our results are not spurious.

Therefore, using the ClustOfVar package in the R repository developed by Chavent *et al.* (2011), the ClustOfVar algorithm helps us to identify groups of variables that exhibit maximum similarity, providing the principal components that are used later to re-cluster firms. This re-clustering helps us to confirm whether variable collinearity effects were dismal. Figure 2.10 illustrate a cluster dendrogram, stability of partition plot, and a box plot for the overall data set. The appendices illustrate the same for 2006 and 2013, in figures 2J and 2K respectively. These are combined to determine the optimal number of clusters of the variables.

In Figure 2.10, the stability plot is the plot of the mean (over B = 1320 bootstrap samples) of the adjusted rand indices. It suggests 3, 4, 6, or 7 clusters. The box plot illustrating the dispersion of these indices over the B = 1320 bootstrap replications for partition suggests either 3, 4, 6 or 7 clusters. The box plot actually projects all solutions as possible. However, the dendrogram projects a 4-cluster solution to be less feasible than the rest. For consistency, the choice of 3 cluster solution is selected.

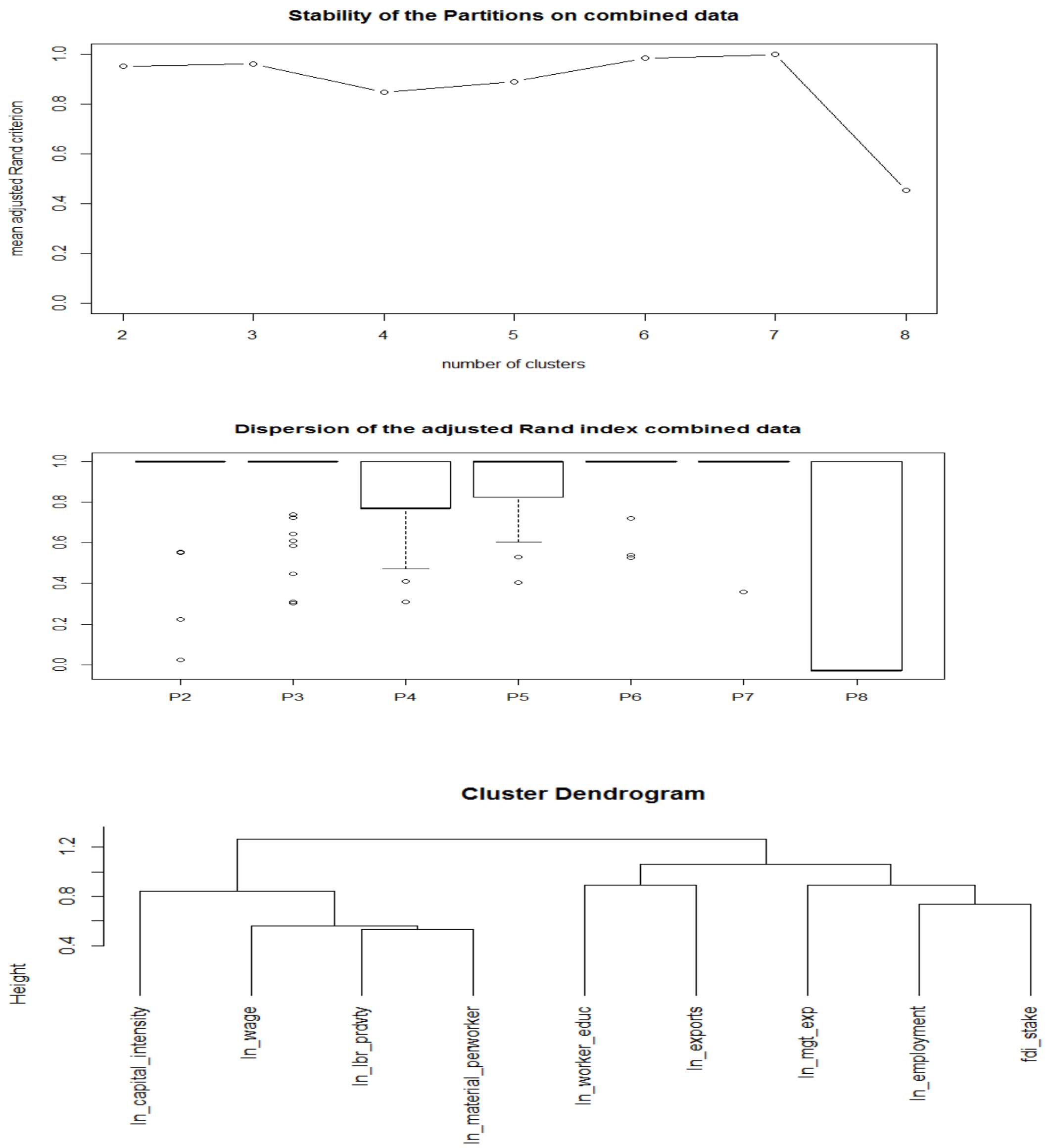


Figure 2. 10: Stability plot, box plot and dendrogram for overall data variable clusters

Additionally, the dendrogram in Figure 2.10 reflects the matrix results. For instance, among all variables, employment posts the highest correlation with fdi_stake at 0.27 and from the dendrogram, employment is closer (clustered) to fdi_stake than any other variable. This result is suggestive of a likely closeness between foreign-ownership and firm size variables. Exports are more correlated with worker_educ (|0.37|) in the matrix and the dendrogram depicts the same. The matrix shows that labour productivity is more correlated with material per worker at 0.71 than with wage, at 0.55 level, an aspect also echoed by the dendrogram in Figure 2.13.

In the appendices, based on 2006 sub data set results, figure 2.J shows the stability of the partitions plot for the mean (over B = 500 bootstrapped samples) of the adjusted Rand indices. It suggests 3, 4 or 5 clusters. The box plot illustrating the dispersion of these indices over the B = 500 bootstrap replications for partition suggests either 3 or 4 clusters. A close inspection of the dendrogram highlights 4 clusters being more feasible than 3 clusters. In Figure 2.K, based on the 2013 sub data set results, the stability plot is the plot of the mean (over B = 700 bootstrap samples) of the adjusted Rand indices. It suggests 3 (to a lesser extent) 4, 6, or 7 clusters. A closer look at the box plot illustrating the dispersion of these indices over the B = 700 bootstrap replications for partitions suggests either 3, 4, 6 or 7 clusters. The box plot actually projects 3 clusters to a much similar extent like the 4, 6 and 7 solutions. The dendrogram on the other hand, feasibly reveals 3, 6, or 7 clusters than 4 cluster solution.

Table 2.12 shows the composition of the clusters of variables for each cluster solution on the three sets of data, together with the factor loadings. For each cluster, the squared loadings of each variable with the synthetic variable central to the cluster is noteworthy. The synthetic variable is the first principal component of PCAMIX. These squared loadings correspond to the squared correlation ratio for the variables, with the synthetic variable. It should be recalled that the clustering was performed using only numeric variables otherwise the squared loadings would also correspond to the correlation ratio for non-numeric variables in some cases. For example, the squared correlation between the variable "*ln_employment*" and the central synthetic variable of cluster 1 for 2006, 2013 and combined data is 0.64, 0.56, and 0.59 respectively. The correlation ratio for *"ln_wage"* in cluster 2 for all data sets is 0.68, 0.62, and 0.57 respectively for 2006, 2013 and combined data.

Table 2. 12: Synthetic variable cluster composition for respective data sets

2006		2013		Combined Data	
Clusters	squared loading	Clusters	squared loading	Clusters	squared loading
Cluster1		*Cluster1*		*Cluster 1*	
ln_employment	0.6357	ln_employment	0.5649	ln_employment	0.5860
ln_mgt_exp	0.6357	fdi_stake	0.4985	fdi_stake	0.4707
Cluster2		ln_mgt_exp	0.2820	ln_mgt_exp	0.3105
ln_lbr_prdvty	0.8724	*Cluster2*		*Cluster 2*	
ln_material-perworker	0.7739	ln_lbr_prdvty	0.5428	ln_lbr_prdvty	0.6087
ln_wage	0.6796	material_perworker	0.4564	ln_capital_intensity	0.2686
Cluster3		ln_wage	0.6221	material_perworker	0.6132
fdi_stake	1.0000	*Cluster 3*		ln_wage	0.5773
Cluster 4		ln_capital_intensity	0.3513	*Cluster 3*	
ln_capital_intensity	1.0000	ln_worker_educ	0.3406	ln_worker_educ	0.5557
		ln_exports	0.5296	ln_exports	0.5557

Source: Author's own computations based on World Bank Enterprise Survey 06/13 panel

Focusing on the clusters of overall data set, Table 2.12 results show that variable clustering echoes the dendrogram and the correlation matrix. For instance, cluster one houses employment, management experience and fdi stake. These variables are tightly set apart on the dendrogram too. Variables in cluster 2 are visibly separated from the rest as per the dendrogram and indeed are clustered alone while worker_educ and exports join up to form the third cluster.

Specific inspection of cluster 1 (synthetic variable 1) for 2013 and combined data proposes a variable that brings information on ownership and human capital features of the firms. Cluster 2 (synthetic variable 2) for 2013 and combined data is suggestive of a synthetic variable that communicates production technology, efficiency and remuneration. Finally, cluster 3 (synthetic variable 3) tends to suggest a variable concerned with human capital skills and market orientation of firms. The clustering of variables is consistent between 2013 and combined data with only capital intensity clustered in different clusters for the two data sets.

The algorithm yields scores of each firm on the three synthetic variables (cluster of variables). These scores are then used to re-cluster the firms, doing this for both the 2013 and overall data sets. For consistency, a similar methodological procedure is utilised in re-clustering. The dendrograms in Figure 2.11, show that the data sets and associated re-clustering yield three cluster solutions as before. This confirms that earlier results from cluster analysis of firms are not detrimentally affected by variable collinearity, which provides a robustness check for the results.

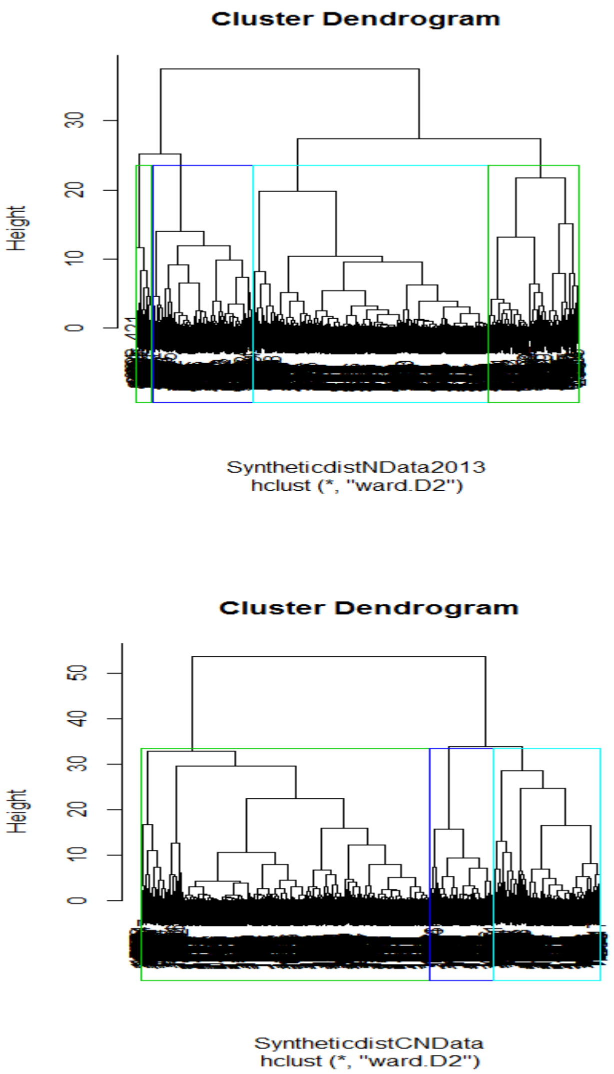


Figure 2. 11: Hierarchical clustering of firms using synthetic variables

This sub-section of variable clustering confirms the presence of multicollinearity amongst the variables used in clustering. Consequently, a dimension reduction procedure is performed on the data so as to generate uncorrelated principal variables. Using these variables to re-cluster the firms yields similar cluster solutions as before. Getting similar cluster solutions shows that despite the presence of multicollinearity, its assumed effects are not significant enough to invalidate the clustering results. This implies that the clusters generated and the associated descriptions are rich enough to support generation of valid and useful conclusions about the hypothesized systematic differences between foreign and domestically-owned firms in Uganda. These clustering results, however, do not tell us whether such differences are more associated with foreign ownership or with other factors that have been found to be correlated with ownership status. Insights into this association were earlier provided by regression analysis. Furtherance of this understanding can be achieved by investigating the relative importance of the numerous factors in the segmentation of firms, especially the variable on ownership status. I accomplish this endeavour by performing a Classification and Regression Tree (CART) analysis. The results of this analysis and the procedures followed are discussed in the section that follows.

2.6.4 Classification and Regression Tree Analysis (CART)

Overall, it is evident that this agglomerative hierarchical method of segmenting Ugandan firms tends to group these firms based on a variety of variables; ownership status (foreign/local), worker_educ, employment, and other variables. It is, however, not clear which of the variables play a relatively big role in firm segmentation although ownership seems to have a strong influence based on the results generated so far. This is another hypothetical thought that elicits my interest in investigating the relative importance of the variables in firm segmentation and specifically confirming whether a firm's ownership status leads the way ahead of other variables. In this attempt, we perform a CART analysis on the data.

CART is a classification method that uses historical data to construct what literature refers to as decision trees, which are then used to classify new data (Timofeev & Hardle, 2004). It consists of four basic steps; the first is tree building using recursive splitting of nodes, the second is stopping the tree building process when a "maximal" tree has been built, third step is tree "pruning", and lastly optimal tree selection (Lewis, 2000). In its simplest form, we let t_p be a parent node and $t_l$and t_r respectively left and right child nodes of parent node t_p. Consider the learning sample with variable matrix X with M number of variables x_j and *N* observations. Let class vector *Y* consist of *N* observations with total amount of *K* classes. The maximum tree will be constructed based on a splitting rule that maximises homogeneity of the child nodes. This homogeneity is defined by an impurity function. The splitting rule is based on a Gini Index whose algorithm solves the maximisation problem below;

$$x_j \leq x_j^R, j = 1 \ldots M^{\left[- \sum_{k=1}^{k} p^2(k|t_p) + P_L \sum_{k=1}^{k} p^2(k|t_l) + P_r \sum_{k=1}^{k} p^2(k|t_r)\right]} \quad (2.11)$$

Using the overall data set, we develop regression trees and the most important variables are identified. We use *rpart*, an R package that generates comprehensive results of CART, including variable importance based on percentage contribution in the firm segmentation. We further use the "*rpart.plot*" package to generate an illustrative CART tree for visual inspection of how firm segmentation is done. Using the *rpart* package, we standardise the variables first and then run CART regression model of the formula;

as.factor (Cluster) ~ ln_employment + ln_lbr_prdvty + fdi_stake + ln_capital_intensity + ln _material_perworker + ln_mgt_exp + ln_wage + ln_worker_educ + ln_exports_absolute

In this formulation, the dependent variable, cluster, is a categorical variable taken by the R package as a factor. In Table 2.13 we see the relative variable importance in segmenting firms based on overall data set.

Table 2. 13: Variable importance in firm segmentation

Variables	*absolute contribution*	*percentage*
fdi_stake	285.52	75.0
ln_worker_educ	49.89	13.0
ln_mgt_exp	14.63	4.0
ln_wage	9.37	2.0
ln_lbr_prdvty	8.28	2.0
ln_material_perworker	5.02	1.0
ln_employment	4.50	1.0
ln_capital_intensity	1.63	1.0
ln_exports	0.00	0.00

Source: Author's own computations based on World Bank Enterprise Survey data 06/13 panel

Noticeably, ownership status (fdi_stake) posts the highest absolute and percentage level of importance in segmenting firms. Besides this variable, worker_educ is revealed as importantly distinct from other variables in classifying firms in the data. Together with fdi stake, these two variables account for 88% contribution in segmenting firms. This result echoes earlier results, notably statistical significance that indicated the lowest p-value for the mean difference between clusters in terms of worker_educ. The next variable, which is twice as important as the remaining ones, is management experience at 4% while the rest follow with very dismal contributions. It is sufficient to note that the export variable is not important in classifying firms in the data. This finding is further intriguing because available literature tends to show that foreign-owned firms are more export-oriented than domestically-owned firms, which would make exportation an important element in separating local and foreign-owned firms.

These results are visually shown in Figure 2.12 with overall data CART tree:

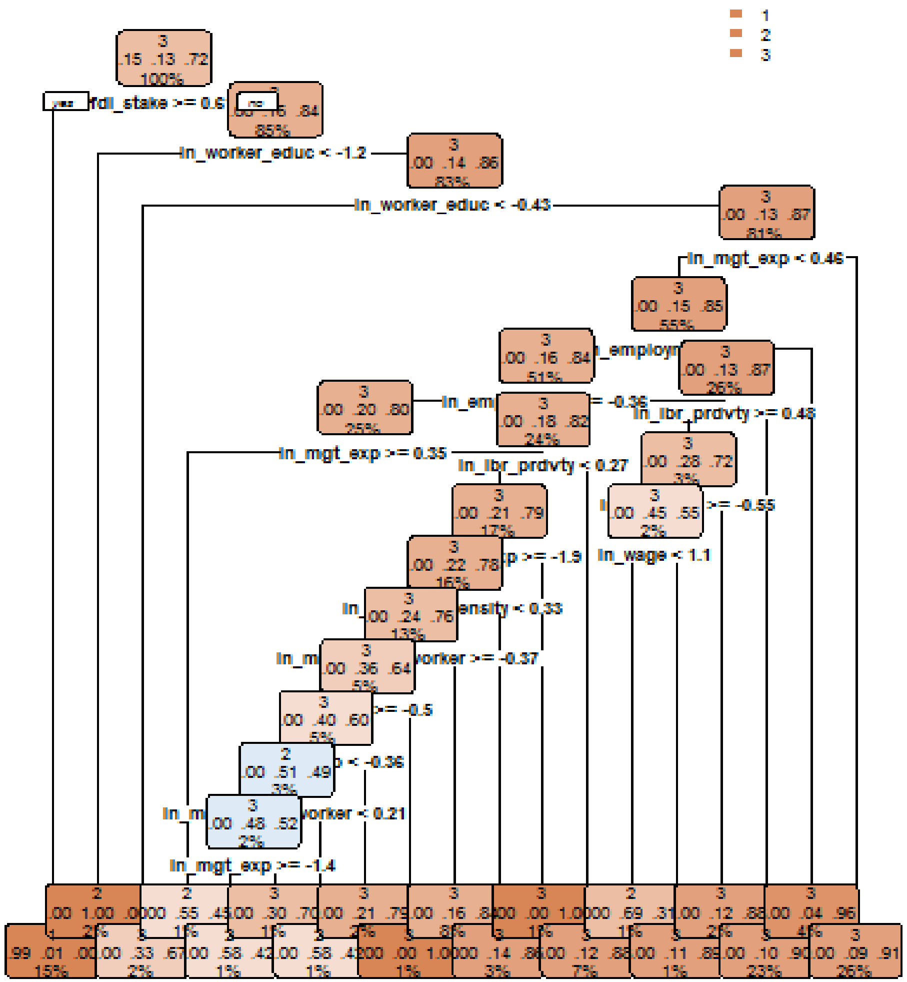


Figure 2. 12: A CART tree diagram based on overall data set

2.7 Conclusions

Earlier studies, mostly conducted in the developed world have established that differences indeed exist between foreign-owned firms and domestically-owned firms on a number of performance indicators. Some of these studies such as Moss and Ramachandran (2004) and Obwona (1998), have been conducted in SSA, specifically the East African region. Most of these earlier studies were based on discriminant functional analytical methods where prior assumptions especially on data distribution dictate the course of investigations. This chapter, with machine learning techniques, uses a relatively novel method, cluster analysis, where the data is allowed to determine the course of investigation to a greater extent. In essence, it provides an equivalent of a robustness check test to earlier findings from other studies on foreign-owned firms' performance characteristics in relation to domestically-owned investments.

In the second analytical stage, standard hierarchical clustering was applied to group the firms using a number of performance and characteristic variables. These variables included: employment, capital_intensity, wages, worker_educ, firm exports, management experience, labour productivity, material use per worker, and ownership status. The analysis grouped firms into largely foreign and domestically-owned. Descriptive statistical analyses of firm clusters revealed that the cluster with largely foreign-owned firms tended to perform relatively better when compared to the cluster with predominantly domestically-owned firms. This was in terms of average real values on the selected performance and characteristic variables mentioned above. Besides finding significant differences between foreign-owned and domestically-owned firms, the later are also found to differ amongst themselves on some indicators, notably workers' education. These results, besides being comparable to findings of available empirical studies, they provide evidence to the inadequately attended to question of; whether observed differences are simply due to comparisons between sub groups of firms or not. Based on these results it can be concluded that foreign-owned firms systematically differ from domestically-owned investments hence the observed performance premiums in favour of foreign-owned firms. These observed differences are not simply a reflection of comparisons between sub groups of firms. By implication then, the observed dismal economic performance outcomes in Uganda amidst increasing inflows of foreign investments could be emanating from other sources whose investigation is beyond this thesis' focus, while actually foreign investments that flow into the country meet the characteristic and performance standards that have been associated with them in the available literature.

Additionally, the chapter establishes foreign ownership's importance in segmenting firms alongside the selected performance indicators in Uganda. It does not, however, solely account for the segmentation. Indeed, after employing ClustOfVar algorithm to cluster variables into groups, findings showed that indeed foreign ownership stake was correlated with a bunch of other variables but more with employment and management experience in firms. Moreover, results from CART analysis equally revealed that ownership status (fdi-stake), worker_educ, and mgt exp were the most outstanding variables in terms of relative importance. Specifically, foreign ownership was more than twice as important as the other variables in influencing the formation of groups of firms. Therefore, in relation to one of the components of the empirical question at hand i.e. whether the observed differences are due to ownership status or not, we can prudently conclude that whereas this might be true, it is only to an extent, with other correlated factors at play. This makes it essential therefore that further investigations are undertaken to disentangle the extents to which the observed systematic performance differences can be ascribed to ownership and other correlated factors. This is an empirical issue attended to in the later chapters of this thesis.

Overall, results in this chapter agree with some empirical components of earlier findings where foreign-owned firms are found to differ from domestic firms in host economies (these differences are also confirmed in Table 2E in appendices, by regression results using clusters as regressands on selected variables). Indeed some studies like Falk and Wolfmayr (2010), Helpman *et al.* (2004), and Melitz (2002) have found that productivity and management know-how of firms contribute to a natural selection of firms that engage in foreign investment.

This exploratory multivariate analysis of data does not provide predictive mechanisms but paves the way for them. By classifying firms into groups that can easily be studied relative to each other, discriminant predictive methods can be employed to study the groups further. Such causal mechanisms are explored further in the chapters that follow, particularly in chapter 4, in which I address the key question regarding the causal effect of foreign ownership on firm-level performance.

One limitation of this analysis is related to data, with missing values being the most visible of all. Due to the low number of firm observations (1325), credible imputations were done to avoid a drop in the number of firms, which would have reduced the data points further.

Additionally, the study used the World Bank Enterprise Survey of 2006/2013, making results subject to economic changes in Uganda over the years. However, in the next few years, a new survey is expected. It would be useful for new research to employ similar techniques on new survey data and possibly other data sets, to generate new results. This new research will ultimately contribute further to better evidence-based policy formulation and implementation.

Chapter 3

Foreign Ownership and Private Firm Enterprises in Developing Countries

Evidence from Sub-Saharan Africa

3.1 Introduction

In Chapter 2 of this thesis, I employed the method of hierarchical cluster analysis to investigate firm-level characteristics of both foreign-owned and domestically-owned firms in Uganda. I specifically investigated whether foreign-owned firms systematically differ from domestically-owned firms along numerous performance dimensions. Using country-specific data for Uganda, I found that indeed foreign-owned firms are likely to systematically differ from domestically-owned firms. Findings indicated that actually observed performance differences between the two types of firms may not be related to just comparisons between sub-groups of firms but these firms differ systematically in terms of structural and performance characteristics. Foreign ownership is found to be a key variable in segmenting firms, but it is also correlated with numerous other variables. These results are to some extent similar to findings from earlier empirical studies. Some earlier empirical studies like (Goswami & Kanta, 2012; Tambunlertchai, 2009; Ullah, Shah & Khan, 2014) have also provided evidence in support of these firm-level features that tend to distinguish the two types of firms and are more likely to be associated with the observed performance differences. Many of these empirical studies and their findings have also been country specific.

One of the implications of country-specific studies is that findings may not easily be applicable to other relatively similar economies. For instance, unlike the findings of other empirical studies like the ones mentioned above, I did not find similar heterogeneities among foreign-owned firms in Uganda i.e. the algorithm did not optimally cluster foreign-owned firms into numerous feasible clusters with significant differences. This slight difference in findings of chapter 2 and other earlier studies may probably be due to, among other reasons, the narrower and country-specific data set utilised in Chapter 2. And therefore, this could be a pointer to uncertainty regarding the applicability of my country specific findings in relatively similar economies, especially in SSA. Moreover, foreign direct investments constitute a global economic phenomenon and findings may vary across countries or regions, even when the same economic phenomenon is investigated.[17] It is unclear therefore, although very important for empirical investigations, whether findings in Chapter 2 are a country-specific phenomenon or they could be generalised across similarly developing economies, especially in SSA. Given that the focus of this thesis is SSA, then in terms of empirical literature furtherance, this chapter investigates a similar empirical question from a multi-country perspective. To add nuance to the anticipated findings, further analysis is provided from the perspective of group classification of economies, based on World Bank/IMF benchmarks i.e. low-income versus lower middle-income economies. The availability of a broader data set covering 19 countries in sub-Saharan Africa (SSA) provided this opportunity. And for matters of consistency, I replicate the same methodology utilised in chapter 2.

The cardinal objective was to establish how common or unique country-specific findings from Chapter 2 are, within a multi-country context, specifically among economies in SSA, the region of focus in this thesis. In relation to the cardinal objective, I additionally aimed to establish whether foreign-owned firms also exhibit heterogeneities among themselves at firm level, as available empirical literature asserts. In so doing, this chapter avails evidence that supports or dispel the hypothetical argument that findings from Uganda are robust, as expected, across relatively similar economies in the developing world, especially SSA. This chapter and its findings, not only gives nuance to previous results but also lends supportive consistency in terms of applicability from one economy to another, at a time when the African continent is envisioning the possibility of a single economic block. In addition, this chapter provides a solid base for the generation of testable hypotheses.

In this chapter, as in Chapter 2, cluster analysis of characteristics of firms in SSA indicates a group of firms that are predominantly domestically-owned and are more likely to differ from other groups of firms on a number of performance dimensions. Relative to other groups, firms in this cluster are generally likely; to be small or medium in size, to exhibit relatively lower labour productivity, to be associated with lower mean wages, to exhibit relatively low levels of incremental investments, and to export part of their output. However, key to this cluster formation is the fact that firms in this group are quite similar hence their being clustered together irrespective of country of origin or whether clustering is performed on income groups of economies classified using World Bank yardsticks. This suggests that differential features

[17] See Gaston & Issouf (2012) on the linkage between FDI and poverty reduction in Africa.

of domestically-owned and foreign-owned firms in SSA are generally the same; and, further, that findings obtained from Chapter 2 based on Uganda are not peculiar and can be generalised to other economies, especially in SSA and other developing economies. This key result was consistently visible when I clustered the data in the following ways: firstly, the entire data set; secondly, the four individual countries with the largest number of sampled firms, excluding Uganda; and thirdly, even when I clustered lower-income and lower-middle-income economies separately.

Additionally, after separation of the cluster composed of largely domestically-owned firms, firms that exhibited foreign ownership were optimally grouped into four different clusters. These clusters tended to differ from each other on several dimensions, yielding the common phenomenon of intra-firm heterogeneity among foreign-owned firms. Such heterogeneity is visible between Clusters 2 and 3, as well as between Clusters 3 and 4, for mean wages, mean employment, capital intensity, exports, labour productivity, worker training and average material use per employee. The phenomenon of firm-level heterogeneity is common in the scholarly findings of several studies on foreign direct investments in which different methods of analysis have been applied.

In this chapter, clustering results tend to suggest that heterogeneities amongst clusters of foreign-owned firms are more likely to be visible in terms of foreign ownership intensity and export intensity, as opposed to country of location or classification in terms of low-income and lower middle-income categories. The cluster that is more likely composed of exclusively foreign-owned firms, is less export intensive when compared to two clusters that are unlikely to constitute exclusively foreign-owned firms. These two clusters were not significantly different from each other except in terms of foreign ownership intensity. The result of firm-level heterogeneity among clusters of foreign-owned firms was also consistent when I clustered both the entire data set and the four individual countries with the largest number of sampled firms excluding Uganda; and even when I clustered lower-income and lower middle-income economies separately. This consistency confirms that findings for Uganda, discussed in Chapter 2 of this thesis, were not exceptional.

The classification and regression tree analysis indicated that ownership status, exportation, sales, labour productivity, employment, wages, capital intensity, and employee training are important variables that sort firms into particular clusters. The remainder of this chapter is laid out as follows:

- Section 3.2 presents empirical literature on private investments in SSA;
- Section 3.3 presents the data description;
- Section 3.4 presents the methods of analysis and results;
- Section 3.5 contains a qualitative analysis of case studies of foreign-owned firms in selected countries in SSA; and
- Section 3.6 presents the study conclusions.

3.2 Stylized facts and empirical literature on foreign and locally-owned investments in SSA

Since the 1990s, the inflow of foreign direct investments in Africa has been on the rise although with a declining global share. In 2016, global flows of foreign direct investments reduced by about 2% to USD 1.75 trillion. Investment in developing countries declined even more – by 14 % – and flows to least developed countries (LDCs) and structurally weak economies have remained volatile and low (UNCTAD, 2017). The majority of economies in SSA and in Africa in general are categorically developing economies; with some definitely being structurally weak and suffering of the decline and volatilities in foreign direct investment flows.

Recent trends show a diminishing African share of global foreign direct investments. For instance, between 2014 and 2016, Africa's foreign direct investment inflows declined from USD 71bn to USD 59bn (UNCTAD, 2017). In SSA, sluggish commodity prices have reduced economic prospects and investor interest in the sub region. One of the leading foreign direct investment recipients on the continent, Angola, registered declines, as did the large economies of South Africa and Nigeria (UNCTAD, 2017). Moderate increases were registered in East Africa, with an increase in inflows going to Ethiopia. However, strong inter- and intra-regional trade agreements are expected to lead to more foreign direct investment inflows in SSA and into Africa in general in the near future.

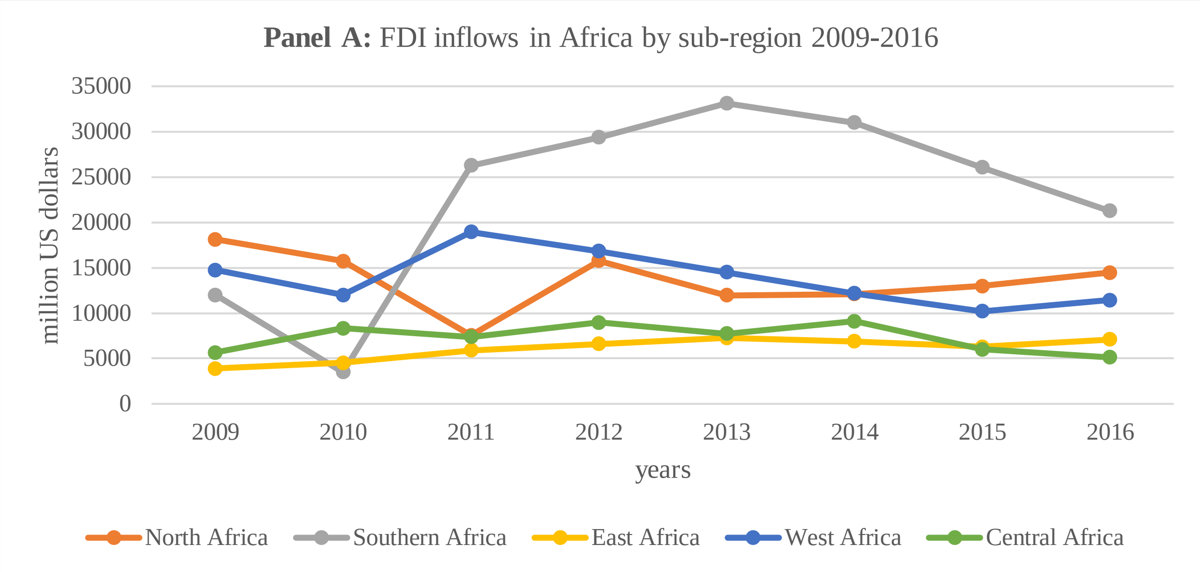
Regionally, West Africa has been taking the lead in the recent past, driven mainly by natural resource discovery in countries like Ghana and Ivory Coast (UNCTAD, 2014). According to UNCTAD (2017), this success has been the result of fiscal consolidation and self-imposed reductions in government investment spending in Ghana. In Ivory Coast, foreign direct investment inflows rose by 17 % to USD 675 million, reflective of supportive public investments by the government as well as economic diversification. In Figure 3.1, Panel A, the regional distribution of foreign direct investment inflows in Africa is shown. The illustration shows that, besides Southern Africa (which is driven by the giant economy of South Africa), the West African region and the North have, on average, been leading the rest of Africa, with the central region performing marginally. With reference to SSA, the Western region leads, followed by the Central and East African regions. East Africa has witnessed more inflows than Central Africa, a point that is clearly highlighted by the graphical illustrations. In Panel B, the trends in foreign direct investment inflows for the countries in the data set are shown. The series shows the averages for the group including Uganda (Group+Ug), for the group excluding Uganda (Group-Ug), and for Uganda alone between 1990 and 2018. It is noticeable that the trends for Uganda's inflows and the mean for the rest of the countries are similar over the years. This relative similarity further justifies the extension of my Chapter 2 analysis to investigate whether findings in Uganda might hold for relatively similar economies in the region, or whether it is a country case phenomenon.

More than any other factor, natural resources are a key factor that attract significant inward foreign direct investment, as illustrated by offshore gas and oil exploration in Mozambique and Ghana, mineral wealth in the DRC, and copper and cobalt in the Congo. Elsewhere, infrastructure and manufacturing has driven foreign direct investments in countries like Ethiopia. Among the peculiar case studies is Rwanda, whose major attraction for foreign direct investments has been rigorous economic reforms and institutional quality (Chen, Geiger & Fu, 2015). Chen *et al.*'s conclusions are supported by earlier case study results on SSA by (Basu & Krishna Srinivasan, 2002).

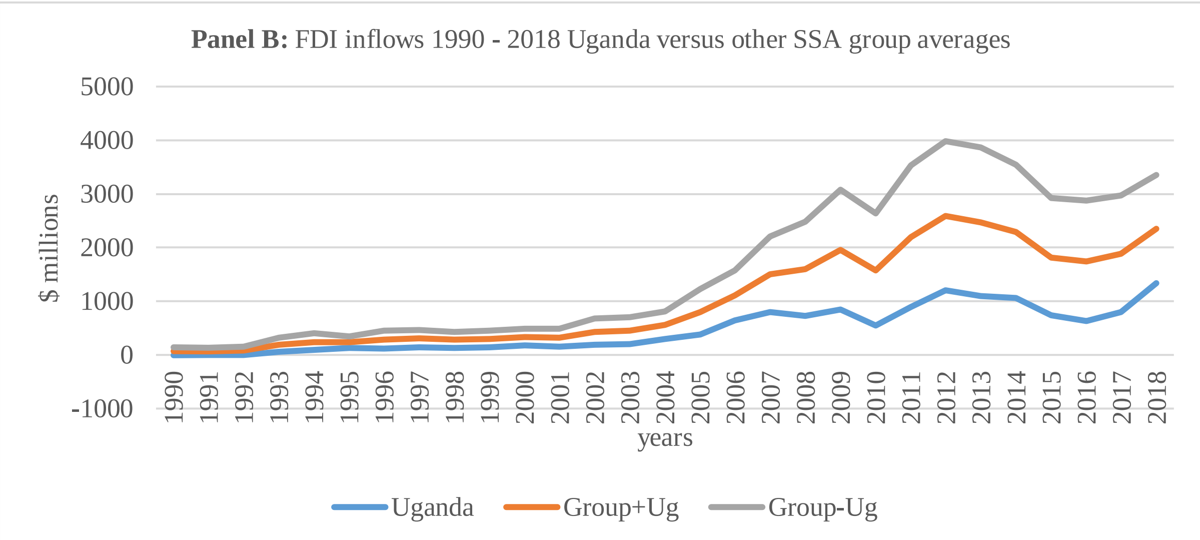
In SSA, foreign direct investments are more prevalent in the manufacturing sector than in other sectors like services, although the latter is slowly increasing. Recent evidence shows that manufacturing foreign direct investment in SSA is 9%, higher than the global average of 7.5% and the developing countries' average of 8.1% (Chen *et al.*, 2015). The agricultural sector has the lowest foreign direct investment inflows, with available data showing that agricultural foreign direct investment is very small when compared with domestic agricultural investment. A review of case studies on SSA indicates that, on average, less than 5% of foreign direct investment goes to agriculture (Gerlach & Liu, 2010: 8).

Foreign direct investment in manufacturing in SSA continues to see some diversification in terms of source and new partners coming on board, especially China, Brazil and India. However, there are still unique cases of sectoral destination of foreign direct investments, with

Rwanda registering more foreign investments in information and communications technology (ICT) and financial services in the last decade.



Source: Author's own illustration based on UNCTAD, FDI-MNE Information System[18]



Source: Author's own illustration based on UNCTAD, FDI-MNE Information System

Figure 3. 1: African foreign investment inflows, by sub-region, and Uganda in comparison to other SSA countries (US $millions)

Besides the stylized facts discussed above, some scholarly investigations about private investments in SSA have been conducted although these still remain few probably due to data

[18] FDI database at www.unctad.org/fdistatistics

limitations. Some of these studies attempt to answer the very empirical questions posed in this thesis although utilizing conventional methods. Empirical firm-level literature on SSA includes studies investigating the features of both foreign-owned firms and domestically-owned firms. One of the focus areas of such studies has been the salient differences between foreign-owned and domestically-owned firms. Empirical findings on characteristic differences between foreign-owned and domestically-owned firms in SSA have indicated that the former are relatively better performers than the latter, and that they consequently differ systematically along various variable dimensions. Coniglio *et al.* (2015) employ OLS and MM estimation methods to estimate a CES production function for a sample of 6497 firms in SSA. Their findings indicate that foreign-owned firms for instance pay higher (16.9%) wages on average than domestically-owned firms do, *ceteris paribus*. Utilising the same data set as Coniglio *et al*, Blanas *et al.'s* (2017: 17) empirical findings reveal that foreign-owned firms pay a mean wage that is 31.9% higher than the mean wage paid by domestically-owned firms. These results are in line with empirical findings based on economies elsewhere, such as the findings of Demurger and Fournier (2005) in China. This wage premium has been found to be a disincentive for cross-firm labour mobility, which ultimately impedes knowledge diffusion from occurring between foreign-owned and domestically-owned firms[19]. The wage premium has been attributed to a number of factors in the literature. It may be explained by unobservable worker characteristics, such as higher ability or greater motivation (Javorcik, 2015), or by the tendencies of foreign-owned firms to operate in high-wage sectors and high-wage locations (Moran, Alfaro & Javorcik, 2007: 28). This explanation is comparable to Almeida's cherry-picking argument. In some host economies, the concentration of foreign direct investments in high wage sectors may be a consequence of regulatory frameworks in place. In Ethiopia, for instance, several sectors are not open to foreign direct investment; however, the majority of these sectors are those that feature low performance characteristics like low wages. Examples are restaurants, retail and coffee exportation. This indirectly implies that most foreign direct investments in Ethiopia end up in sectors with wage premiums. This further implies that ceteris paribus, findings in chapter 2 of this thesis may not hold for an economy like Ethiopia since the observed performance premiums in favour of foreign-owned firms may be due to such firms being anchored in sectors that were already characterised by high; wages, output, capital/labour ratios, productivity, technology levels, among other performance outcomes.

[19] See studies by Poole (2013) and Balsvik (2011).

SSA firm-level empirical literature further indicates that foreign-owned firms are different from domestically-owned firms in terms of size and skill intensity (Coniglio *et al.*, 2015: 1256). This means foreign-owned firms in SSA tend to employ more workers or create more jobs compared to domestically-owned firms. The conclusion regarding size by Coniglio *et al*, echoes findings by Chen *et al* (2015) who equally established that manufacturing foreign direct investments in SSA are associated with more employment creation (hence firm size) than in any other sector, providing evidence from case study countries like Tanzania and Uganda in 2011 and 2012 respectively. Moreover, employment in foreign-owned firms tends to be more stable and secure than the employment offered by domestically-owned firms in SSA (Blanas *et al.*, 2017). Employment instability in domestically-owned firms might be due to low survival rates among such firms due to their low levels of resilience to most shocks. Higher job security and stability in foreign-owned firms might be attributed to the need by such investments to ensure that their foreign affiliates undertake critical operations (such as the production of intermediates and final output for supply to foreign markets) in line with the demands of parent firm headquarters, probably to maintain brand parity. It may also point to better human resource management practices associated with foreign-owned firms since they tend to employ skilled managers, as Chapter 2 results indicated.

Foreign and domestically-owned firms in SSA have also been found to be different in terms of export intensity, capital intensity, and labour productivity, with domestically-owned firms found to be less intensive along most dimensions. Njikam (2018) investigates the linkage between export-market destination and firm performance in SSA. Besides finding superior characteristics of exporters relative to non-exporters, his preliminary findings indicated that foreign ownership is key in enhancing the propensity to export, hence it is not surprising that foreign-owned firms are found to be more export-intensive than domestically-owned firms. Rankin *et al.* (2005: 14) find a similar result, being the significance of foreign ownership on exporting, in their analysis of firms in Kenya, Ghana, Tanzania, Nigeria, and South Africa.

Additionally, empirical firm-level literature shows that rising levels of capital intensity are associated with increases in age and foreign ownership of enterprises in SSA (Njikam, 2018: 10). The association between capital intensity and age might be attributed to risk management practises implying that initially foreign investors would not want to commit enormous resources in a foreign country before getting well acquainted with the practical economic environment. Relatedly, regarding ownership stake, it could be that with time as foreign investors gain confidence in say a partially acquired formerly domestic firm and the general economic environment, they acquire more stake which most likely comes with more investment levels in terms of assets.

In the same study, increases in firm size and foreign ownership are found to be associated with improvements in the labour productivity of SSA enterprises (Njikam, 2018: 11). Foreign-owned firms are both export- and capital-intensive, probably due to the demands emanating from international market standards. In addition, these two features are usually complemented by high levels of skill intensity, as exporting firms usually face international competition with dimensions that are different from domestic markets.

Almost universally, empirical findings have found that foreign ownership is associated with higher wages, higher productivity, and increased wage inequality due to increases in skill premium between domestically-owned firms and foreign-owned firms (Hale & Mingzhi Xu, 2016: 2). Majority of the studies, however, remain focused on the developed world as indicated in chapter 2 with little investigations in SSA. In our current context, the few studies that have focused on SSA have majorly been case studies with only few analyses based SSA as a whole hence the generalizability gap of country-specific studies still remains.

Foreign-owned firms in SSA have been found to not only systematically differ from domestically-owned firms but are also heterogeneous amongst themselves. Besides finding differences between foreign and domestically-owned firms, the empirical analysis of Coniglio *et al* (2015) reveals firm-level heterogeneities among foreign-owned firms in SSA. Substantial differences are found between foreign investors from developed and developing economies in terms of both skill intensity and wage premiums. Chinese firms have for instance been found to employ more workers (mostly blue-collar) workers) and to pay lower wages for both skilled and unskilled workers when compared to both domestic firms (-23.7%) and other foreign investors (-49.8%) (Coniglio *et al.*, 2015: 1260). These results partly concur with the empirical findings of a case study analysis of Chinese foreign direct investments in SSA by (Kaplinsky & Morris, 2009). Though an important firm-level feature in most empirical studies which have actually guided policies aimed at incentivising foreign investment inflows, findings in chapter 2 do not reveal such heterogeneities probably due to reasons earlier explained. On another note, heterogeneities among foreign investments may have varying implications in terms of anticipated effects not only on locally owned firms but also the general host economy as well. Therefore, the dismal economic outcomes observed amidst increased inflows of foreign investments in SSA might be due to the nature of most foreign investment inflows in the period under consideration. For instance, are they export oriented (hence more export intensive than local firms) or domestic oriented (hence the dismal export performance that is observed)?

As if to confirm the above implications, empirical studies have also found foreign direct investments to be heterogeneous in their impact on host economies and on domestically-owned firms. Pfeiffer, Görg & Perez-Villar (2014) utilise the World Bank Enterprise Survey data sets to analyse the horizontal productivity effects of foreign-owned investments from industrialised and developing economies in 10 SSA countries. Their findings indicate that the productivity effects of foreign direct investments are heterogeneous, hugely driven by the income levels of host economies and the absorptive capacities of locally-owned firms. South-to-South foreign direct investments[20] are found to be slightly more advantageous than North-to-South[21] foreign direct investments (Pfeiffer *et al.*, 2014: 26).

Concisely, empirical findings on private firm enterprises in SSA clearly suggest that foreign-owned firms are different from domestically-owned firms in terms of firm-level characteristics. Foreign-owned firms in SSA tend to be associated with relatively superior performance outcomes. Besides these differences, intra-firm heterogeneity has also been empirically investigated and confirmed in SSA among foreign-owned firms. Most of the scholarly investigations, as expected, have employed conventional econometric methods to investigate the above differences. Analyses using these methods are based on well set prior assumptions, especially regarding distribution. There has been less or completely no application of alternative methods like cluster analysis that does not impose assumptions on data used but lets the data itself classify firms. This chapter employs this relatively novel method on a broader firm-level data set covering a sizeable number of economies in SSA to investigate the systematic differences alluded to in the literature. The ultimate aim is to confirm earlier empirical findings using alternative methods of analysis and establish applicability of findings in the previous chapter to SSA.

3.3 The Data

This chapter utilises the Africa Investment Survey (AIS) data, 2010, collected by the United Nations Industrial Development Organization (UNIDO). The data consists of over 6000 firms, which were all surveyed in 2010 in the various countries. During data collection, face-to-face

[20] Foreign investment inflows from developing economies to fellow developing economies.
[21] Foreign investment inflows from developed to developing economies.

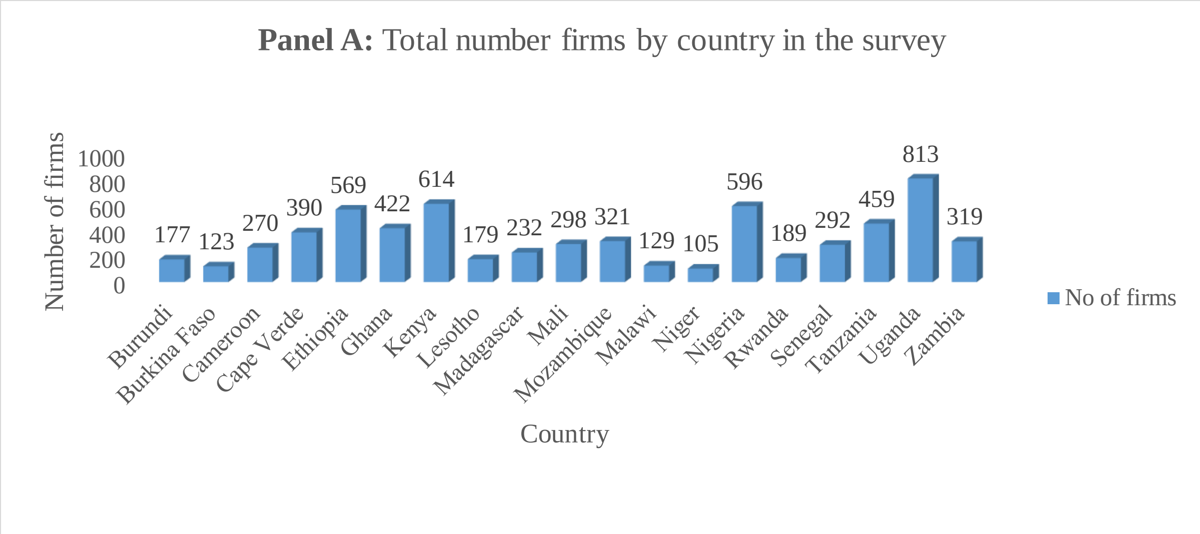
interviews were conducted with top-level managers of foreign-owned and domestically-owned firms. This is akin to how the ES data used in the previous chapter is collected. AIS 2010 covered 19 countries in SSA: Burkina Faso, Burundi, Cameroon, Cape Verde, Ethiopia, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Niger, Nigeria, Rwanda, Senegal, Tanzania, Uganda and Zambia. These countries jointly accounted for 42% of the total GDP of SSA, making the data set a fair representation of SSA. The AIS data collection process is based on a rigorous survey methodology for strategic sampling and interview techniques, making it an authoritative data set for scholarly research and for analyses of foreign and domestic investment in Africa.

Similar to the World Bank Enterprise Survey data used in chapter 2, sampling followed stratified sampling on three dimensions i.e. sector, size and ownership and the primary sampling unit of the AIS was also the firm. Stratified sampling as earlier noted in the previous chapter, has numerous merits notably the fact that it increases the likelihood of precise estimates. The AIS survey covers domestic and foreign owned firms in manufacturing and service sectors just like the ES data utilised in Chapter 2. The manufacturing sector covers sub-sectors like agriculture, forestry and fishing, mining and quarrying (including oil & gas), construction, electricity, gas and water supply, and all other types of manufacturing in metal and non-metal, plastics and rubber. The service sector includes sub-sectors like; wholesale and retail, hotels and restaurants, transportation and storage, tourism, IT and telecommunication, insurance and banking, real estate, consultancy, and education and health, among others.

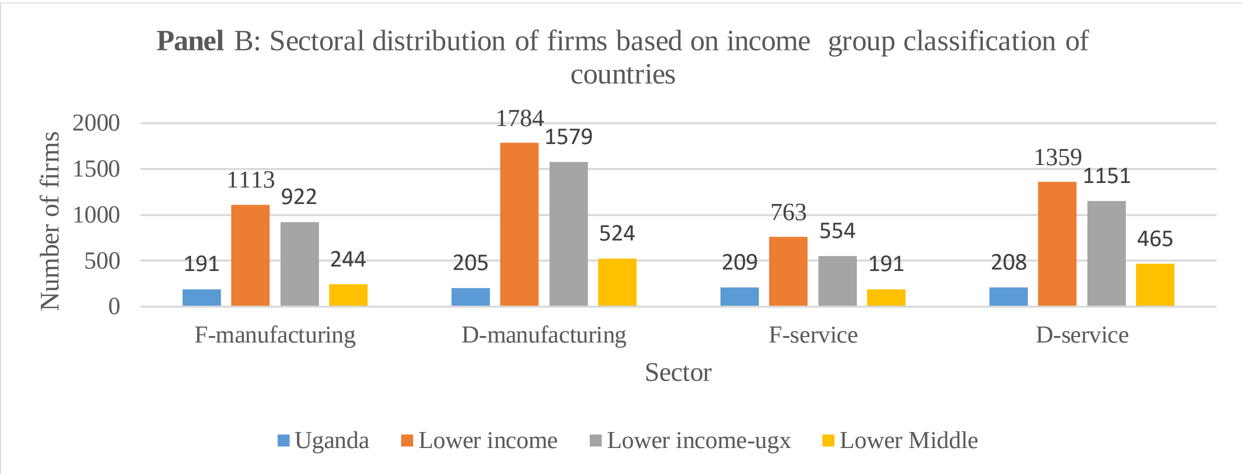
The survey covers various variables at firm level, but categorically ranging from investor characteristics like organizational structure, share structure, interactions with other firms and market orientation. Also covered is financial information of firms like output and production factors, such as labour, capital, energy, intermediate goods, wages and aspects regarding international trade. These are, to a large extent, the variables also covered by the ES data. I selected only those variables deemed relevant for my study analysis. A brief descriptive summary of the data set follows.

In Figure 3.2, Panel A, the distribution of firms for the 19 countries in the sample is shown. This reveals that Uganda has the highest number (813) of sampled firms, followed by Kenya (614), Nigeria (596), Ethiopia (569) and Tanzania (459) each having over 450 firms in the sample. The huge sample composition of Ugandan firms lends more credence to the cardinal objective of this chapter. Burkina Faso, Burundi, Lesotho, Malawi, Niger and Rwanda have fewer than 200 firms sampled each. With the exception of Lesotho, they are low-income economies, which may explain the low numbers of firms. This is also a likely indicator of the smaller size of their private sector.

On Panel B, the distribution of firms by sector and ownership status in two regional World Bank income groupings is shown. The following can be noted: Uganda has an almost-balanced representation in all; the lower income group excluding Uganda (Lower income-ugx) posts higher figures in both manufacturing and service compared to the lower middle-income group; and the same is so for both foreign and domestic firms. This is due to the small number of lower middle-income economies in the sample. Overall, there are more domestically-owned firms than foreign-owned firms, irrespective of groupings and sector. In all the groupings, there are more firms in manufacturing than in the service sector. On Panel C, the net inflows of foreign direct investments, measured as a percentage of GDP for each country in the sample in 2009 is shown. Madagascar posts the highest values, followed closely by Ghana, Mozambique and Cape Verde.



Source: Author's illustration based on AIS-UNIDO data 2010



Source: Author's illustration based on AIS-UNIDO data 2010

Figure 3. 2: Distribution of firms by country, regional income grouping by sector and fdi stocks by country

3.3.1 Variable selection, measurement, and transformation

Theoretically, this analysis was guided by the ownership advantages theory and Dunning's eclectic theory, which were well detailed in chapter 2 of this thesis. This was because most of the variables that are used in the analysis are those covered by these two theories in their explanation of the foreign investment firm in a typical host economy environment. Therefore, selection of variables used in the analysis was largely guided by these two theories. However, earlier studies on Uganda, such as Obwona (1998), and Wilson and Cacho (2007), studies on firms in other economies like Naughtin and Rankin (2016) who employ similar methods also guided my selection variables. Many scholars, in their analysis of firms in other economies on either foreign direct investments or related topics, have used these variables. Studies by Rankin

(2013), and Naughtin and Rankin (2016) on firms in the South African economy, along with Forte and Santos's (2015) cluster analysis of foreign direct investment in Latin America are some examples using these variables. Moreover, some scholars propose that, in clustering, common sense also plays a role. I selected a mixture of variables, being those thought to be associated with firm performance. I also selected some that are purely characteristic in nature, for example, exportation, capital intensity, and ownership status. These variables, including their labels and how they are measured, are presented in Table 3.1. However, I also generated other variables as and when deemed necessary during the analysis.

Table 3. 1: Variable Description

Variable name	Variable label	Variable Measurement
Employment	lnemployment	log (№ of employees)
Labour productivity	lnlbr_prdvty	log (real firm output/№ of employees)
Foreign ownership stake	fdi_stake	% of foreign owned firm shares
Capital intensity	lncapital_intensity	log (NBV machinery/№ of employees)
Material per worker	lnmaterial_perworker	log (real material value/№ of employees)
Wage	lnwage	log (labour cost/№ of employees)
Firm exports	lnexports	log (real firm exports)
Firm output	lnsales	log (real firm sales)
Employee training	lnemployee_training	log (employee training)
Research investment	lnR&D	log (research investment)
Initial investment	lninIv	log (initial investment)
New investment	lnnewIv	log (new investment)

Source: Author's own representation based on AIS, 2010 data set

Given that variables measured in monetary terms were captured in respective countries' local currencies, I converted all monetary variables to US dollars using each country's foreign exchange rate as at 31 December 2009[22]. Additionally, I performed a log transformation on the variables in order to remove any inherent skewness and to make patterns in the data more interpretable. After this transformation, I generated descriptive statistics. Variables that were

[22] Information on GDP deflators, exchange rates, GNI per capita, and group thresholds were extracted from https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?end=2016&start=2010

selected for clustering further underwent a linear transformation. These variables were standardised so that all had a mean value of 0 and a variance of 1. This standardisation was performed because data usually consists of variables which are measured using different scales, and this usually makes comparison challenging. Also, centering variables is well-suited to cluster analysis. As previously, in this chapter I defined a foreign-owned firm as one with a minimum of 10% ownership shares belonging to individuals or parties of nationalities other than the country in question. I dealt with missing values in the same way as applied in Chapter 2.

3.3.2 Creation and Use of Indicator Variables

In order to understand whether the conclusions reached in Chapter 2 were country (Uganda) specific or a phenomenon in SSA (and probably similar for developing economies elsewhere), I created some indicator variables and used them in the description and further analysis of clusters. Based on the World Bank classification of economies, I created a variable indicating whether a country was a low-income economy or lower middle-income economy. This was because all the countries in the survey belonged to one of these two classifications. I also created a variable that indicated a specific country, running from 1 to 19. This helped me to identify whether Ugandan firms were uniquely clustered or not. The expectation was that, if Ugandan firms were unique from the rest in SSA or even other relatively similar economies elsewhere, the algorithm should have clustered them together, taking into account the performance variables used in the segmentation process. By implication, the subsequent CART analysis should have broken the cluster(s) away from the rest.

3.4 Study Analysis, Methods and Results

In my endeavour to further understand the characteristics of foreign and domestic firms; and whether the observed systematic differences between foreign-owned and domestically-owned firms across various dimensions, as documented in Chapter 2, were exclusive to Ugandan firms or were a SSA (or developing economies' or groupings within) phenomenon, I conducted two stages of analysis.

The first stage involved descriptive statistics using means and medians. I generated statistics for the whole SSA data set, for Uganda, and for the two income groupings based on World Bank classification. Within these groupings, I separately generated statistics for foreign and domestic firms. This preliminary descriptive statistical analysis helped me to: (i) compare foreign and domestic firms within groups and across country groupings; and, (ii) get insights regarding properties of the distributions of study variables, for instance, the property of distributional normality.

I also ran a regression using firms' ownership status (fdi stake) as the regressor and selected variables as regressands and controls. Some of the regressors used include labour productivity, capital intensity, employment, exports and wage. Using firm ownership status as a dummy equal to 1 if the firm was foreign-owned and 0 otherwise, I obtained insights into how foreign ownership is likely to be associated with aspects such as firm size, productivity and wages in SSA. The regression results also helped me to shed light on preliminary differences between foreign-owned and domestically-owned firms. For a specific firm i in the sample, I estimated a simple regression model based on the specification in (3.0):

$$ln\, y_i = \gamma_i + \psi f d_i + \xi X_i + \varepsilon_i \qquad (3.0)$$

Where $ln\, y_i$ denotes the regressand of interest for instance employment, γ_i is the intercept, and fd_i is a dummy, which is equal 1 if the firm is foreign owned and 0 otherwise. ψ is the coefficient of interest and X_i is a vector of control variables of interest, for example, sector or country of location, and ε_i is the disturbance term. I estimated these regressions using an R package, *robust*, found in the CRAN repository. The subsection that follows presents the results of descriptive statistics and regression.

3.4.1 Descriptive Statistical Results

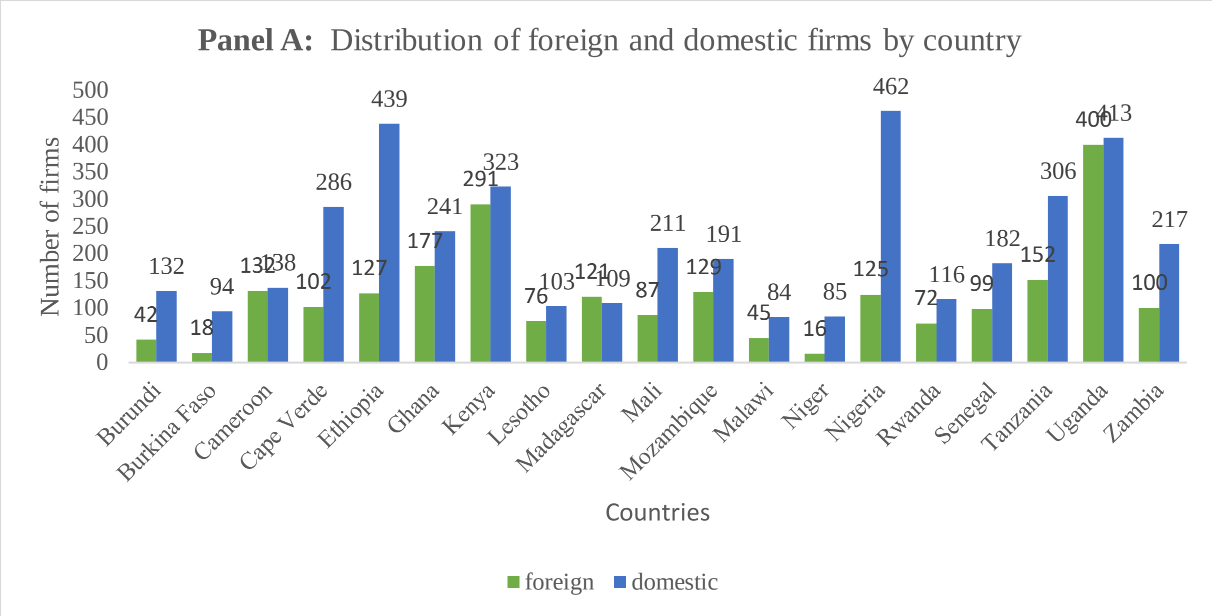
In Figure 3.3 Panel A, the distribution of firms by country for both foreign-owned and domestically-owned firms is shown. Here, it is noted that Uganda has the highest number of foreign-owned firms, followed by Kenya. Uganda having the highest number of firms in the sample is good for comparative purposes with results in the previous chapter. Niger, Burundi, Burkina Faso, and Malawi have the least number of foreign-owned firms in the sample. In all the countries in the sample, as probably expected, domestically-owned firms are more than foreign-owned firms in each country.

In Panel B, the stocks of foreign direct investment per country in the survey in 2009 are shown. According to OECD (2018)[23], foreign direct investment stocks measure the total level of direct investment at a given point in time, usually the end of a quarter or of a year. The outward foreign direct investment stock is the value of the resident investors' equity in. and net loans

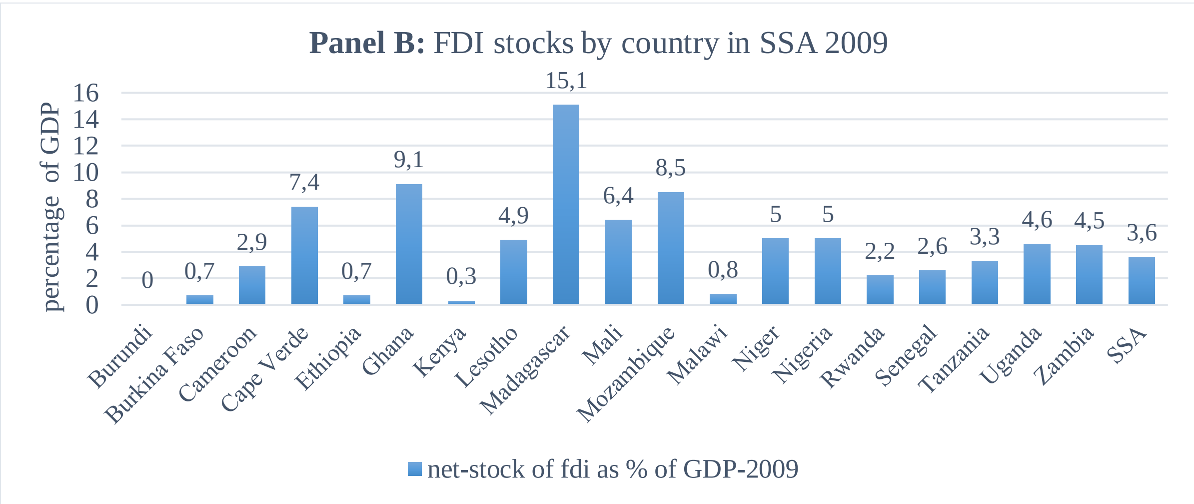
[23] https://data.oecd.org/fdi/fdi-stocks.htm.

to, enterprises in foreign economies. The inward foreign direct investment stock is the value of foreign investors' equity in, and net loans to, enterprises resident in the reporting economy. Foreign direct investment stocks are measured in USD and are shown as a share of GDP[24]. It is noticeable that Madagascar had the highest stocks of foreign direct investment in 2009, (albeit with lower numbers of sampled firms) followed by Ghana, Cape Verde, Mozambique and Mali. A close look at Uganda in Panel B reveals that its foreign direct investment stock was among the lowest even when though it had a balance of domestically-owned firms and foreign-owned firms in the sample.

[24] OECD (2018), FDI stocks (indicator). doi:10.1787/80eca1f9-en (Accessed on 17 August 2018).



Source: Author's illustration based on AIS-UNIDO data 2010



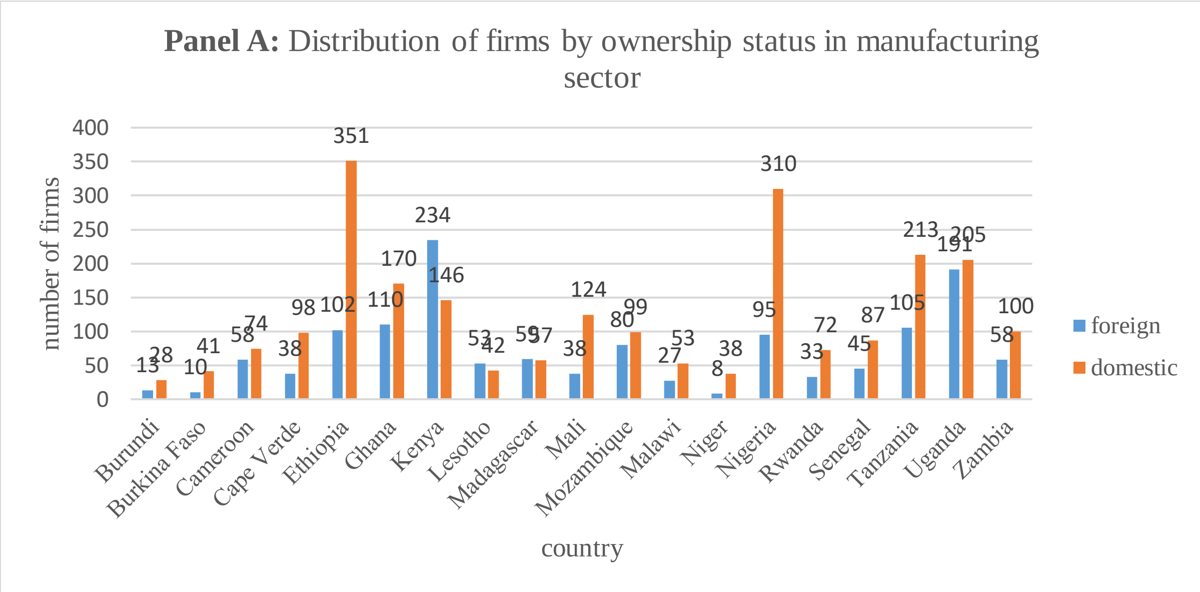
Source: Author's illustration based on data extracted from the World Bank data sets[25]

Figure 3. 3: Foreign direct investment stocks and distribution of foreign & local firms by country

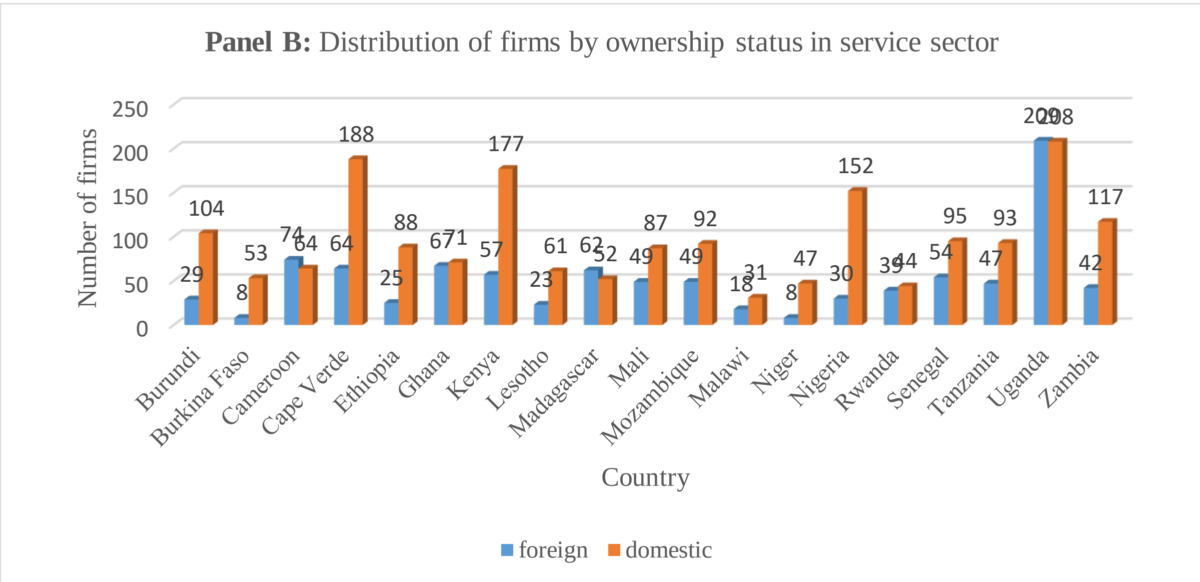
In terms of sectoral distribution, foreign-owned firms tend to be more concentrated in manufacturing over service sectors. The reverse, however, may not hold for domestically-owned firms in most countries. In Figure 3.4 Panel A, the distribution of firms in the manufacturing sector per country is shown. In Panel B, the number of firms for the service

[25] https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS?end=2017&start=2009&year_low_desc=false

sector are shown. From Panel A, it is shown that there are 234 Kenyan foreign-owned firms; however, there are only 25 in Panel B. The same results are shown for, Kenya, Tanzania and Zambia. However, Uganda presents a mixed distribution.



Source: Author's illustration based on AIS-UNIDO data 2010



Source: Author's illustration based on AIS-UNIDO data 2010

Figure 3. 4: Distribution of firms by sector and ownership status in sampled countries

Overall, the sectoral composition of foreign-owned firms reveals that most are concentrated in heavy industries, while domestically-owned firms are mainly found in light manufacturing and services. This sectoral composition of foreign-owned firms tends to support the hypothetical belief that foreign direct investment has the efficacy to stimulate structural transformation of economies of developing countries through creating employment in modern sectors. However, if foreign direct investment is directed to ICT and financial services, job creation and thus anticipated transformation may not be realised. The case study of Rwanda, where average growth of employment is still low amidst increasing foreign direct investments (Chen *et al.*, 2015: 33), is an example of this.

In Table 3.2, the regression results of some selected performance variables and characteristics of firms against foreign ownership and some control variables are shown. For each response variable or characteristic, I first used fdi_dummy as a lone regressor; and, in the alternative estimations, I controlled for the sector, the income group classification of economies, and on whether the firm was exporting or not. I specifically controlled for whether the firm was in the service or manufacturing sector (sector), whether the firm was in a low-income or lower middle-income country (country_class), and whether the firm was exporting or not (export_dummy). Model estimations are shown as M1 to M8, with each regressand corresponding to two models; one with only fdi_dummy, and the second with the three control variables. The results show that foreign ownership is significantly and positively correlated with mean wages, firm size as mirrored by the employment variable, labour productivity and capital intensity. These results hold even after controlling for sector, exportation, and the income group of the economies in the sample. The regression coefficients, which indicate such association, are all positive and significant at least at 1% level. For instance, in models M1, M3, M5, and M7, foreign-owned firms tend to pay more wages, employ more, exhibit more labour productivity, and are more capital intensive by respectively 15.1%, 56.7%, 25.7% and 20.7%, when compared to domestically-owned firms. In summary, these are the likely average differences between foreign-owned firms and domestically-owned firms along the abovementioned dimensions. The coefficients remain significant even after controlling for sector, export status and income group. Additionally, the fdi_dummy coefficient remains positive and significant even when I controlled for the specific country of location of the firms[26]. This implies that foreign-owned investments or firms are likely to pay, on average, higher wages; to employ more workers (hence being larger in size); exhibit higher labour productivity; and be more capital-intensive than the locally owned firms. They are equally

[26]I have not included these results in the table but can supply them upon request.

likely to be more export intensive. An empirical causal relationship between foreign ownership and exporting is popular in current literature, for instance in Rankin *et al.* (2005: 13).

The association between foreign ownership and key firm-level variables as indicated by the regression results above, shows that foreign-owned firms are likely to be different from domestically-owned investments. They, however, do not indicate whatsoever, whether foreign-owned firms also differ among themselves. The need to fill these gaps led me to the next stage of analysis where I employed clustering techniques.

Table 3. 2: Linear regression results of firm performance variables and foreign ownership

Variable	*log wage*		*log employment*		*log labor_productivity*		*log capital intensity*	
	M1	M2	M3	M4	M5	M6	M7	M8
fdi-dummy	0.151**	0.110*	0.567***	0.393***	0.257***	0.185**	0.207***	0.173**
	(0.055)	(0.057)	(0.036)	(0.036)	(0.085)	(0.090)	(0.077)	(0.081)
Sector1		0.106*		0.383***		-0.190		0.935***
		(0.056)		(0.035)		(1.607)		(0.082)
Country_class(1)		0.993***		-0.236***		1.256***		0.346***
		(0.060)		(0.042)		(0.105)		(0.086)
Export_dummy		0.179***		0.809***		0.417***		-0.063
		(0.066)		(0.042)		(0.093)		(0.090)
Constant	2.620***	2.291***	3.830***	3.542***	3.920***	3.755***	3.166***	2.516***
	(0.033)	(0.050)	(0.021)	(0.031)	(0.053)	(1.607)	(0.047)	(0.077)
Observations	4078	3798	6400	5774	3066	2880	3089	2900
R^2	0.002	0.069	0.038	0.143	0.003	0.053	0.002	0.048

*Note: *p<0.1; **p<0.05; ***p<0.01,*
Values in parenthesis are standard errors
Source: Author's own results based on UNIDO's AIS Data, 2010

In the appendices, Table 3A1 shows further descriptive statistics. These statistics show significant differences between foreign-owned and domestically-owned firms. The p-values associated with t-tests on group means for the variables confirm how the two types of firms differ significantly. Overall mean and median values for all variables are nearly similar. This approximate similarity is suggestive of a symmetric distribution exhibited by the variables. In Table 3A2, also in the appendix, descriptive statistics for foreign and domestic firms grouped for Uganda; low-income countries excluding Uganda; and lower middle-income countries, are shown. Each grouping has a column for foreign (F) and domestic (D) firms. Noticeable in the results are visible differences between mean and median real values for Uganda and the other two groups. Mean values for Ugandan firms are lower even when over 12% of the firms in the data are sampled from Uganda. Additionally, firms in lower middle-income economies posted higher mean values than both low-income economies and Uganda. This may be due to differences in firms' attributes in the two categories of economies, with firms in slightly wealthier economies performing better. Finally, differences between foreign-owned firms and domestically-owned firms for the three groups are also evident, with foreign-owned firms having higher values than domestically-owned firms.

The first stage of analysis provided insights similar to the key result in Chapter 2. However, these descriptive statistics did not reliably confirm whether differences between foreign and domestic firms were attributable to comparisons between different sub-groups (for example between large firms where foreign direct investment is more prevalent and small firms) or to systematic differences across various dimensions between foreign and domestic firms. In particular, they did not answer whether the results

(i) were unique for Uganda, as Chapter 2 findings suggested; or

(ii) showed features that characterised private firms in SSA and developing economies, and hence were not specific to Uganda only; or

(iii) were applicable to other country groupings such as the World Bank classification of economies, for instance, low-income economies compared to lower middle-income economies.

I further investigated the above questions in the second stage of this study's analysis.

3.4.2 Standard Hierarchical Clustering of SSA firms

Clustering is the second stage of analysis in this chapter. Methodologically, and unlike in Chapter 2, I selected ten variables, being seven performance variables and three characteristic variables. Using the hot deck imputation method, as described in Chapter 2, I addressed missing values. Next, I standardised these variables to a mean and variance of 0 and 1 respectively, since they were measured on different scales. After this transformation, I undertook a Cluster Tendency Assessment (CTA) using the Hopkin's statistic to establish if the data set was clusterable. It is not credible to subject a data set without naturally inherent clusters to any kind of partitioning (Banerjee & Davé, 2004). The Hopkin's statistic and the five steps it entails are as described in Chapter 2. After confirming the clusterability of the SSA dataset, I investigated whether the variables were free from extreme levels of collinearity.

As noted in Chapter 2, even if the data used contains naturally inherent structures, when variables used in clustering are highly collinear, there is a likelihood that results from the clustering process might be of less quality or sometimes spurious. Collinearity among variables is highly likely when clustering is based on Euclidean distance measures. In this study, I employed Gower's distance measure as the first mitigation measure against any likely spuriousness arising from collinearity of clustering variables. Secondly, I tested the quality of the results by clustering variables to generate principal components (synthetic variables), which I used to re-cluster firms and confirm whether clustering results exhibited consistency at firm level. Unlike in regression analysis where unreasonable beta coefficients can, for instance, intimate the presence of the problem of variable collinearity, in clustering this is difficult due to the absence of betas and dependent variables. In order to detect the presence or otherwise of high variable collinearity, I generated a correlation matrix for the variables that had been selected for use in the clustering. This matrix helped to shed light on the probable presence of extreme collinearity; and hence justifying mitigation actions to be taken or otherwise.

After confirmation of clusterability of the data using the H statistic and ascertaining whether clustering variables were free from high levels of collinearity, I employed confidently hierarchical clustering methods using agglomerative techniques. This was because of the many advantages associated with these techniques, for instance the flexibility in terms of granularity levels and the versatility (Abbas, 2008). I used Gower's general similarity coefficient, which is specified in equation 2.4, to estimate the distances between firms. Based on these distances, the similarity between firms was achieved, this forming the basis for hierarchically clustering these firms. The agglomerative method begins with each firm as an individual cluster, then later links the most similar firms into a new cluster. The process continues until one cluster is achieved. Several linkage methods are used in cluster analysis. Many scholars (Everitt *et al.*, 2011; Řezanková, 2014; Yim & Ramdeen, 2015) have assessed these methods, noting their strengths and weaknesses. In this chapter, I linked the firms using Ward's minimum variance method, a linkage method which minimises the sum of squares to form clusters.

In this study, three methods were used to guide in determining the optimal number of clusters to analyse. These are the dendrogram, the silhouette method[27] and the CH index. The dendrogram is a graphical illustration of how clusters of firms are formed; and by cutting it at a relatively larger distance, an analyst can visually determine the desired number of clusters. Horizontal lines drawn at specific distances on the Y axis of dendrograms indicate the cutting points, where every vertical branch of the dendrogram crossed by the horizontal line indicates a cluster. The number of such crossings give the analyst visual insights into the optimal number of clusters. Elsewhere, the R software enables the analyst to use boxes to indicate such clusters for visual inspection.

The silhouette width is an index associated with Rousseeuw (1987). It is a composite index that reflects the compactness and separation of clusters. For a given firm *i*, its silhouette width S_w is specified in (3.1):

$$s_w = \frac{\lambda(i) - \phi(i)}{\max\{\phi(i), \lambda(i)\}} \tag{3.1}$$

where $\phi(i)$ is the mean distance of firm *i* to other firms in the same cluster, $\lambda(i)$ is the mean distance of firm *i* to firms in its nearest neighbour cluster. The mean of S_w across all firms shows the overall quality of the clustering result. A larger averaged silhouette width (ASW) shows a superior overall quality of the clustering result. In summary, an ASW of 0.71–1.0 indicates that a strong structure has been found, while an ASW of 0.51–0.70 indicates that a reasonable structure has been found. An ASW of 0.26–0.50, however, shows that the structure is weak and that it could be artificial, with a recommendation that additional methods of analysis should be explored. An ASW of less than 0.25 shows that no substantial structure has been found and that cluster analysis techniques are inapplicable. As detailed in Chapter 2, the third method guide, the CH index, developed by Calinski and Harabasz (1974), is based on a variance ratio criterion as specified in equation 2.9.

[27] Generated through the use of package Clustersim in the R software.

Once the clusters had been generated, I comprehensively analysed them using descriptive statistics along the various variable dimensions. The analysis was based on the first sample moment and its focus was on:

(i) The comparison of cluster mean values to the overall mean values of the entire sample along performance dimensions;
(ii) The comparison of individual cluster mean values relative to other clusters in the sample for both performance and characteristic variables selected for analysis; and
(iii) The statistical testing on the means of variables between individual clusters to confirm statistical significance of inherent differences if any. In this endeavour, I used student T- tests.

In executing the three methods described above, this study focused attention on the indicator variables used earlier, namely country, low-income in comparison to lower middle-income economies, sector dimensions and any classifications elements deemed important in understanding firm-level features in SSA. Additionally, keen interest was given to how clusters with foreign-owned firms were likely to compare with those clusters with domestically-owned firms and those clusters with fellow foreign-owned firms, if any. These specific elements of analysis, coupled with the last stage of analysis (CART) were expected to answer the two cardinal objectives in the affirmative. After applying the above clustering process, the results that follow were generated.

The analysis at this stage began with a statistical assessment of the clustering tendency on the dataset using the Hopkin's statistic (H). The H statistic measures the likelihood that a given dataset is generated by a uniform data distribution. It is based on five key steps as detailed in Chapter 2. Using factorextra[28], a Hopkins test statistic of 0.003 was generated, as indicated in Table 3D in the appendices. This statistic is well below 0.5; and on this basis, the null hypothesis that my data was uniformly distributed and therefore did not contain meaningful clusters was rejected.

In Table 3.3, the correlation matrix for SSA is shown. It is evident that majority of the variables in the data are highly correlated. For instance, lnlbr_prdvty is highly and significantly correlated with almost all variables except fdi_stake. Given these correlation results, I

[28] An r package for assessing clustering tendency and for determining optimal cluster numbers using silhouette, Gap statistic and Elbow methods.

concluded that my clustering variables did suffer from extreme collinearity problems, which warranted mitigating actions. Therefore, in this chapter, I also undertook a validation test for the reliability of the clustering results. Methods similar to those employed in the previous chapter, i.e. clustering the variables themselves, so as to generate principal components, were adopted. These principal components were later used to re-cluster firms to establish whether a similar quality cluster solution (5 clusters) could be achieved. Through this re-clustering process an analyst is able to confirm whether firm clusters earlier generated are consistent and reliable.

Table 3. 3: Correlation Matrix for SSA Data Set

Variables	1	2	3	4	5	6	7	8	9	10
1. lnemployment	1.00	.191	.141	.095	.072	.483	.489	.352	.590	.163
2. lnlbr_prdvty		1.00	.731	.917	.782	.742	.651	.643	.889	.009
3. lncapital_intensity			1.00	.639	.572	.537	.519	.510	.636	.004
4. lnmaterial/worker				1.00	.673	.620	.666	.508	.801	-.000
5. lnwage					1.00	.577	.445	.587	.673	.010
6. lnexports						1.00	.529	.638	.874	.076
7. lnresearch_investment							1.00	.693	.745	.032
8. lnemployee_training								1.00	.732	.001
9. lnsales									1.00	.139
10.fdi_stake										1.00

Source: Author's own results based on AIS Survey Data, 2010

The rejection of the null hypothesis based on the H statistic paved the way for the clustering process using hierarchical agglomerative techniques. Using Gower's distance measure, a distance matrix for the firms was generated and thereafter, these firms were linked using Ward's "ward.D2" linkage method. This yielded a large number of clusters, given that over 6000 firms constituted the sample. It is challenging to attempt any analysis of all the clusters; hence the need to decide on the optimal number of clusters to analyse. In deciding on the optimal number of clusters to analyse, I was guided by three methods namely; the dendrogram, silhouette method, and the CH index.

In Figure 3.5, a visual inspection of the SSA dendrogram suggests between two and five feasible clusters. In Figure 3.6, the silhouette plot for SSA is shown. On the plot, the highest

ASW, which is close to 1.0, is associated with a two-cluster solution. According to the silhouette method, an ASW of 1.0 is the highest mark of clustering excellence. The result in Figure 3.6 seems to suggest that the excellent clustering quality is that one, which yields foreign and domestic firms into two distinct clusters in this case. However, the choice of a two-cluster solution, despite its associated excellent cluster quality, may yield only the clear-cut separation between foreign and locally owned firms. Given the objectives of this chapter, the two-cluster solution makes it challenging to:

(i) Understand whether there is a likelihood of Ugandan firms being grouped in a specific cluster. Yet this would have been confirmation for the cardinal objective of this chapter. I could only confirm that findings in the previous chapter were peculiar to Uganda if and only if the findings in this chapter indicated that Ugandan firms were more likely to be clustered uniquely in a specific cluster or clusters.

(ii) Gain insights into the probable inherent heterogeneity between firms, and whether such heterogeneity was more likely to exist between foreign-owned firms, between domestically owned firms or both. Heterogeneity among foreign-owned firms is a firm-level characteristic common in scholarly findings of most available studies that have employed different methodologies (other than cluster analysis) in analysing foreign direct investment characteristics. It is therefore enriching, in scholarly terms, to establish whether any form of heterogeneity could also be revealed via clustering methods.

Any cluster solution above two clusters is, however, associated with a drastic decline in cluster quality. In a three-cluster solution, for instance, there is a noticeable sharp reduction in ASW, from almost 1.0 to below 0.75. Nevertheless, according to the standard measure mentioned earlier, an ASW of 0.5 or above indicates a reasonable quality of clusters that permits good analysis. It is evident from the silhouette plot in Figure 3.6, that any number of clusters from three and above are associated with an ASW of close to 0.6. On this basis, when I ran the CH index as my last test to determine the optimal cluster number, I set the minimum and maximum cluster-solution between three and five clusters respectively. As seen in Table 3D in the appendices, the CH index suggested a five-cluster solution, which I then used as the basis of my analysis for the overall SSA data set.

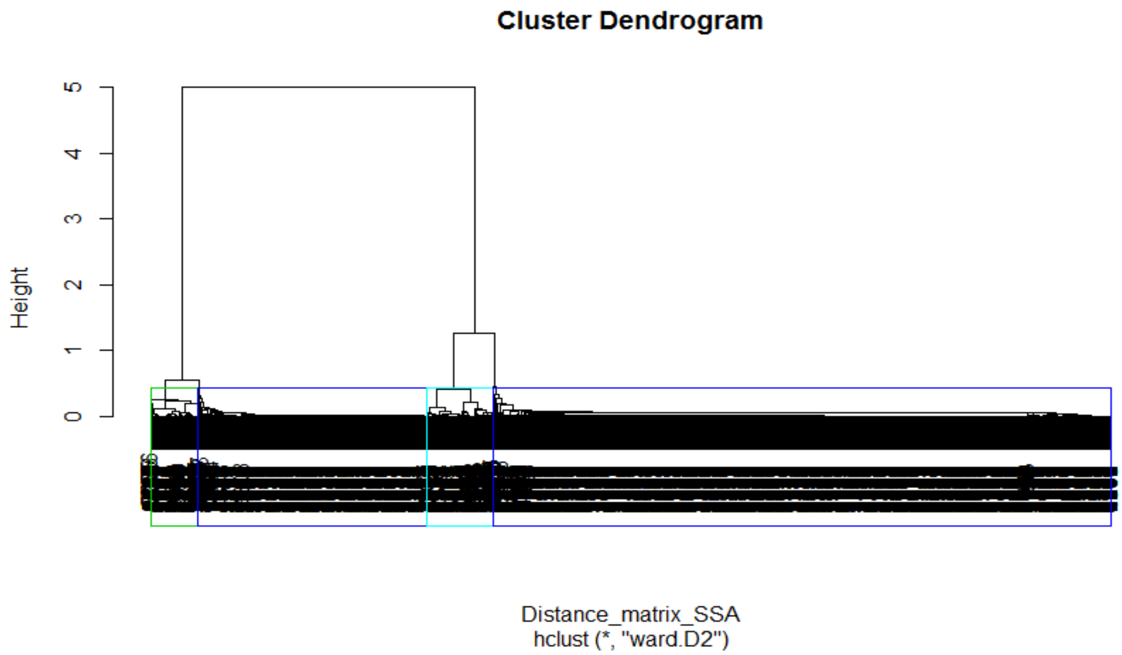


Figure 3. 5: SSA Cluster Dendrogram

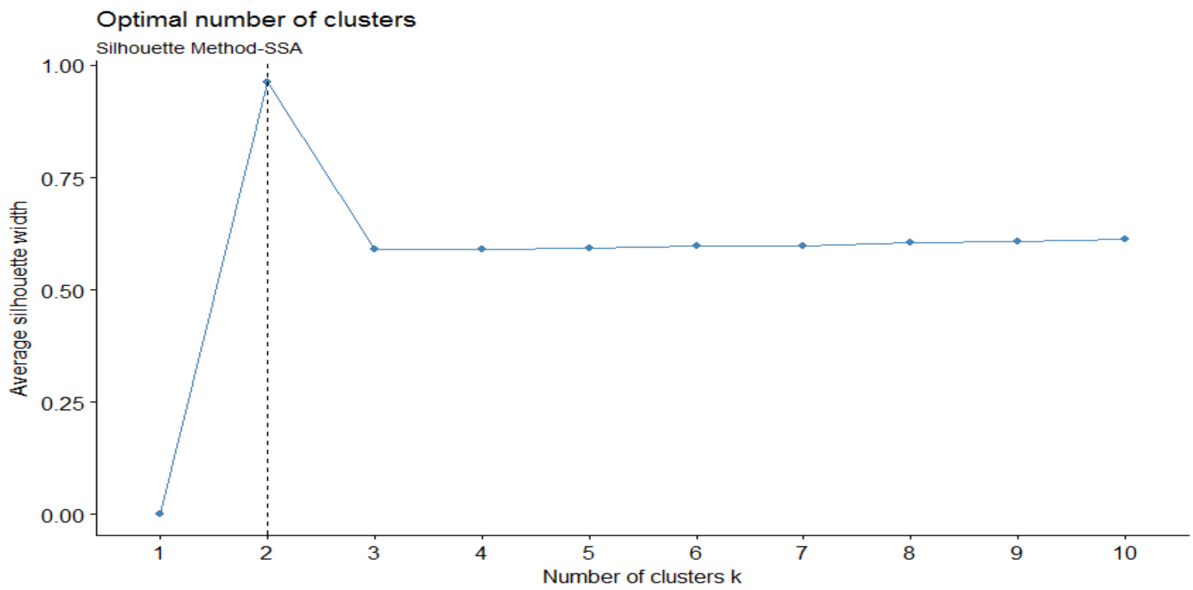


Figure 3. 6: Silhouette Plot for SSA

I decided to cut the dendrogram at a five-cluster solution and to undertake a descriptive statistical analysis of this. Table 3.4 shows the real mean values for each of the five clusters based on the nine variables used in clustering and other variables not used in clustering but deemed informative. These variables are shown as wage (lnw), foreign ownership stake (fdistk), employment (lnempt), capital intensity (lncapint), sales (lnsales), labour productivity (lnlbrpvty), material per worker (lnmat), employee training (lntrn), exports (lnexp), initial investment (lninIv), and new investment (lnnewIv). For each cluster, Table 3.4 further shows the percentage contribution from Uganda (Ug), low-income countries excluding Uganda (LY) and lower middle-income countries (LM). Additionally, shown is the percentage contribution per cluster of both foreign (F) and domestic (D) firms, and also the total number of firms for each cluster in the last column. In the very last row, the mean values for the whole SSA dataset are shown. In this chapter's analysis, I compare values for each cluster to those of the whole SSA data set and also to other clusters. A comparative analytical discussion of each cluster follows.

Table 3. 4: Real mean values and country/group percentage composition for SSA clusters

	Variables											***Uganda & group % composition/cluster***					
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lninIv*	*lnnewIv*	Ug	LY	LM	F	D	**Total**
Mean																	
Cluster 1	2.62	0.01	3.83	3.17	7.37	3.92	3.56	3.41	7.12	6.06	5.95	50.7	65.2	69.7	0.1	99.95	**4184**
Cluster 2	3.07	80.3	4.58	3.76	9.05	4.60	4.18	4.09	8.17	0.79	6.98	3.6	5.3	4.0	13.5	0.0	**313**
Cluster 3	2.59	99.9	4.29	3.08	8.27	3.87	3.45	3.32	7.65	1.15	6.00	41.8	22.3	18.5	67.1	0.0	**1552**
Cluster 4	3.10	41.7	4.62	3.93	9.31	4.83	4.24	4.09	8.39	0.87	7.24	3.9	7.1	7.8	19.2	0.0	**444**
Cluster 5	1.79	45.5	4.10	3.06	17.9	13.0	3.63	-	4.94	6.36	2.99	0.0	0.1	0.0	0.1	0.05	**4**
SSA	2.67	30.6	4.03	3.24	7.83	4.02	3.61	3.49	7.55	5.44	6.11	12.5	65.4	22.1	35.6	64.4	**6497**

Source: Author's own results based on AIS Survey Data, 2010

Cluster 1- 64.4% of the sample (domestic, medium, lowly productive and moderately capital-intensive firms)

Cluster 1 is largely composed of domestic firms (99.5% of all domestic firms in the data set) with only two (0.1%) of all foreign-owned firms in the data set grouped in this cluster. Results in Table 3.4 show, for this cluster, a mean value of fdi-stake of 0.01%, which is below the SSA average of 30.6% in the last row. These largely domestically-owned firms post real mean values on majority variables, which are either below or equal to SSA averages. For instance, they are likely to exhibit lower levels of labour productivity as shown by the cluster's mean value of 3.92, which is below the SSA average of 4.02. They also show low levels of employment. They are also likely to be moderately capital-intensive and medium-sized, as indicated by results on Table 3.4. Available literature shows that these are some of the central characteristics exhibited by domestically-owned firms, especially in developing economies. For instance, a case study of Ethiopia, where most of the significant employment opportunities in manufacturing are attributed to non-local firms, corroborates these findings (Chen *et al.*, 2015: 36).

Comparatively, except along the variable that measures the initial investment level of the firm, (labelled '*lninIv*'), Cluster 1, relative to other clusters, is associated with lower real mean values on all variables. Clusters 2, 3, and 4 are composed of only foreign-owned firms, with mean fdi-stake above the SSA average. In Table 3C in the appendices, the *p*-values associated with test statistics on differences between the mean values of the clusters are shown. These results indicate that differences between the largely domestically-owned firms in Cluster 1 and the foreign-owned firms in the rest of the clusters are statistically significant. These results echo those in Chapter 2, and provide further evidence reinforcing earlier study findings on firm-level characteristics of both foreign-owned and domestically-owned firms in developing economies. Using regression analysis, Coniglio *et al.* (2015: 1249) finds similar results between domestic and foreign-owned firms in sales, labour productivity, export intensity, capital intensity, and mean wages. Elsewhere, Foster-McGregor, Isaksson and Kaulich (2013) use first-order stochastic dominance methods and quintile regressions to investigate the effect of foreign ownership on labour market outcomes in SSA. Their findings also suggest that foreign-owned firms pay between 9.3% and 20.9% higher wages than domestic firms when taking into account mean wages of all employees (Foster-McGregor *et al.*, 2013: 12). Additional findings on employment show that foreign-owned firms employ between 10% and 13% more workers than local firms. Moreover, the employment of blue-collar workers is correspondingly found to be between the same range, with foreign-owned firms employing between 8% and 10% more white-collar workers (Foster-McGregor *et al.*, 2013: 19).

The systematic differences between Cluster 1 (largely domestically-owned firms) and Clusters 2,3 and 4 (largely foreign-owned firms) were also revealed consistently when, in a bid to perform robustness checks on the results, I additionally applied the clustering algorithm on:

- Firms from four individual economies with large numbers of sampled firms but excluding Uganda;
- Firms from low-income economies and lower middle-income economies separately; and
- Firms using all other variables except firm ownership status (fdi_stake).

The clustering results for these three segments of analysis are comparable to findings from the main results. On Tables 3F1 to 3I2, shown in the appendices, it is evident that results from the four individual country analyses are relatively similar to the main clustering results. In each country analysed, Cluster 1 is composed mainly of domestically-owned firms. Firms in this cluster are more likely to be different from the rest of the clusters along most variables, the only exception being some outlying clusters. Specifically, firms in cluster 1 for each country are predominantly domestic and the average values for the cluster along the selected performance variables are lower than those of other clusters that contain largely foreign-owned firms. These country-specific results also confirm findings from chapter 1 which were based on Uganda.

On Tables 3J1 and 3K1 in the appendices, clustering results for low-income economies and lower middle-income economies respectively are shown. In both tables, Cluster 1, which is composed of mainly domestically-owned firms, is associated with real mean values that are either lower or equal to those of the overall group averages which are shown in the last rows of the tables. When compared to the rest of the clusters that are composed of largely foreign-owned firms, cluster 1, in both tables, posts lower mean values on all performance variables. Statistical tests on the differences of group averages between cluster 1 and the rest of the clusters indicate that the observed differences are statistically significant. The *p-values* that confirm this result are indicated in Tables 3J2 and 3K2 in the appendices for respectively, low-income and lower middle-income economies. This implies that even with country groupings based on national income status, foreign-owned firms are likely to be clustered away from groups consisting of largely domestic firms. And the associated average performances are higher in favour of clusters with predominantly foreign-owned firms.

In the third segment, I eliminate ownership stake (fdi-stake) from the list of clustering variables and re-cluster firms again using the overall data set. The aim was to confirm whether even without this characteristic variable, the algorithm can ably cluster away most domestically-owned firms in a particular cluster. Once the clusters are revealed, I compute their average fdi-stake to identify which clusters constitute largely domestically-owned firms and those clusters that constitute predominantly foreign-owned firms. Then the average real values for each cluster along the selected performance indicators are computed and compared with those of other clusters and the averages for the whole group. In Table 3L1 in the appendices, results for the four major clusters of SSA without ownership status as a clustering variable are shown. The algorithm separates away Cluster 1 with a mean fdi stake of 29%, which is still lower than the SSA average of 30.6%. A fdi stake of 29% shows that firms in this cluster are largely domestic and the few foreign-owned that might be in this cluster are largely similar to domestically-owned firms, which is not surprising. Indeed, in this cluster, majority (66%) of the firms are domestically-owned. This cluster systematically differs from the other three clusters (who's mean fdi stake is above SSA average) along all performance variables. Firms in this cluster have, on average, lower real mean values for all variables when compared to firms in the rest of the clusters. Table 3L2, in the appendices shows *p*-values that reveal statistically significant differences between the real mean values of cluster 1 and the rest of the other clusters along the selected performance indicators. Put in another way, these results indicate that clusters with largely domestically-owned firms or with lower mean foreign ownership are likely to differ systematically from those clusters with higher levels of fdi stake on average. This result is reflective of the main results obtained when the algorithm was applied on all firms with ownership status as a clustering variable. This is also an indicator of robustness of our findings

At all four levels of analysis, the clustering algorithm consistently yields clusters of firms, which, when analysed, indicate systematic differences in performance along selected variables between largely domestically-owned firms in one cluster and largely foreign-owned firms in other clusters. This consistently echoes previous findings in Chapter 2 where a narrow, country-specific data set was used. These results prove that observed performance differences between foreign-owned firms and domestically-owned firms are not simply a reflection of comparisons between sub-groups of firms, a feature that holds for SSA as well. It therefore implies that the dismal economic performance observed amidst increased foreign investment inflows in SSA is probably due to other factors largely.

A key implication of the clustering results without the fdi-stake variable is that the ownership characteristic (fdi_stake) tends to influence segmentation more than other variables. However, even without this variable, largely domestically-owned firms are clustered together (66%), which is significant enough to lend credence to our initial results. The importance of this variable in firm segmentation is investigated later in the chapter using CART and for a similar justification detailed in chapter 2 earlier.

A further essential result to note is that all domestically-owned firms from Uganda are clustered in cluster 1, as are other domestically-owned firms from both low-income and lower middle-income economies. This implies that domestically-owned firms from Uganda and other economies in SSA are relatively homogenous in terms of firm-level characteristics, which systematically tends to separate them from foreign-owned firms. Otherwise, Ugandan firms and firms from any other economy in the sample would have been uniquely separated and hence clustered away by the algorithm. In Table 3.4, we notice from the (last six) columns for percentage composition for Uganda and the two World Bank groupings, that in Cluster 1, 50.7%, 65.2% and 69.7% are firms from Uganda, low-income countries excluding Uganda, and lower middle-income countries respectively. This composition further confirms this study's earlier hypothesis that domestically-owned firms in Uganda are unlikely to be unique in any way from those in other relatively similar developing economies. This result partly answers the fundamental question of this chapter i.e. whether the findings in Chapter 2 about Uganda are unique or quite generalizable to other relatively similar economies in SSA and around the world. The other implication here could be that the slight differences observed in chapter 2 results could be related to data, especially its narrowness.

3.3.3 Intra-foreign-owned firms' classification

The clustering algorithm in this chapter does not only reveal a separation (clustering away) of domestically-owned firms (Cluster 1) from foreign-owned enterprises, as discussed above. The algorithm also yields three other clusters, which are composed of foreign-owned firms only[29]. As noted earlier, the three clusters of foreign-owned firms do not only differ from Cluster 1 but also exhibit significant differences between themselves. This is what I alluded to as the phenomenon of intra-foreign-owned firms' heterogeneity earlier in one of the study objectives. These differences therefore demonstrate firm-level heterogeneity, a commonplace

[29] The fourth cluster, bringing the number of clusters to five, is noisy but contains both FDIs and domestic firms.

characteristic associated with foreign direct investments in most scholarly studies that have employed different methodologies and have analysed different data sets. These heterogeneous clusters of foreign-owned firms are presented and discussed below.

Cluster 2 (4.8%) and Cluster 3 (23.8%) – highly and all foreign, large size/capital intensive firms

These two clusters provide the first segment of the intra-foreign-owned firms classification yielded by the clustering algorithm. They both contain only foreign-owned firms but are significantly different from each other on all the variables under consideration, including fdi-stake. This is an indicator of firm-level heterogeneity among foreign-owned firms, which is a commonplace empirical finding in numerous studies that use other methodologies for analysis. On Table 3C in the appendices, the *p-values* associated with statistical tests on the comparisons of real mean values for all the variables between these two clusters are shown. The *p-values* confirm that these two clusters are significantly different (1% level) from each other on all the variables. Significant differences amongst foreign-owned firms might have implications for performance outcomes as well. It is possible that different foreign investments are associated with varying magnitudes of performance and therefore effects on host economies. This finding provides evidence in support of the second objective of this chapter. This evidence is further supported when firms are clustered based on a set of variables excluding foreign ownership. As revealed by results in Tables 3L1 and 3L2 in the appendices, Clusters 2, 3 and 4 are composed of largely foreign-owned firms (with a mean fdi stake above the SSA average) and do indeed differ from one another significantly along the variables considered.

Unexpectedly, in Table 3.4, we notice that Cluster 2 (with a lower fdi-stake on average) is associated with higher real mean values than Cluster 3 (with a higher fdi-stake) on all variables. Yet ideally the expectation is that the more foreign-ownership stake there is in a firm, the more the expected performance, ceteris paribus. Cluster 3 has firms that are more likely to be entirely foreign in ownership and as such they are expected to be associated with relatively higher performance values. This cluster has a mean and a median fdi_stake of 99.9% and 100%, respectively, unlike Cluster 2 which has a mean fdi-stake of 80.3%. The entirely foreign-owned firms are likely to be a result of the complete takeover of domestic firms by foreign investors; or of new investments started in a specific economy. The latter case may be associated with what is popularly referred to as Greenfield investments. The case for firms in Cluster 2 could probably be more associated with mergers and acquisitions (M&As), in which a foreign investor acquires shares in a local firm, instead of setting up a completely new investment project. If these two modes of foreign investment entry are taken as likely possibilities, then the sort of performance contradiction above can be attributed to the 'cherry-picking 'argument associated with Almeida (2007). Foreign investors that usually choose M&As as their entry mode (this being the likely choice we are assuming for firms in Cluster 3 and, most probably, Cluster 4), usually target or cherry-pick those domestic firms that are already performing well. Such performance is usually in labour productivity, wage premiums, and skilled human capital (Georgopoulos & Preusse, 2009: 592; Hale & Mingzhi Xu, 2016: 5).

This rather surprising result can also be attributed to economic policies that spell out the legal ownership requirements for foreign direct investments in some of the countries in the data set. Whereas there is no limitation on foreign equity for any firm in Mozambique, for instance, the reverse is true for Ethiopia. There are several sectors where ownership can only be via a joint venture with domestic investors, with the latter having to own not less than a 27% stake (UNCTAD-I&PR, 2002:27). Moreover, the capital requirements for joint ventures are lower when compared to wholly foreign-owned investments, making joint ventures (M&As) more attractive and more likely than greenfield investments, irrespective of how economically strong the intending foreign direct investor might be. Highly productive foreign direct investments are therefore likely to enter the market through M&As as opposed to Greenfields due to regulations in place, hence the likelihood of M&As (Cluster 2) outperforming Greenfields.

In Zambia, British investors Tate & Lyle, originally started the leading foreign-owned firm in sugar production, Zambia Sugar Plc., in 1960. However, the government's nationalisation program dictated that 51% shares be sold to government in 1972, transforming it into a joint venture. The Ghana Oil Palm Company in Ghana is another leading foreign owned investment in the palm oil industry whose current ownership structure is attributed to government liberalisation policies of the 1990s (Sutton & Kpentey, 2012: 37). In Ethiopia, regulatory restrictions are even more stringent, as shown by the following quotation:

> *"… although the foreign direct investment regulatory framework in Ethiopia is more open now than during the 1980s, it is still highly restrictive compared to many other developing countries, including those in the region"* (UNCTAD-I&PR, 2002:27).

Apart from cherry-picking and regulatory explanations, it is also a fact that some leading foreign-owned firms in SSA economies are more inclined to serve domestic markets. These firms are unlikely to be as productive or sophisticated as their counterparts, which are inclined to serving international markets[30]. In Zambia, one of the leading foreign-owned investments, Hybrid Poultry Farm (Zambia) Ltd, only exports 1–2% of its output (Sutton and Langmead, 2013: 46). Safintra Zambia Ltd, a South African foreign-owned firm in the building materials sub-sector, does not export its output at all. In Ghana, Takoradi Flour Mills, a leading foreign-owned investment in agro-processing, sells all its output domestically (Sutton & Kpentey, 2012: 57). Elsewhere, Coast Millers Ltd., a foreign owned firm in agro-processing in Tanzania and a key player in the industry, only exports 30% of its output (Sutton & Olomi, 2012: 71).

In addition, the surprising result can also be related to the kind of sectors where most foreign direct investments might be found, especially entirely foreign-owned firms. Ghana in West Africa has witnessed high foreign direct investment inflows in mining and oil exploration, with the Global Investment Report 2018 predicting continued foreign direct investment inflows to Ghana. This prediction is based on massive investments by firms like Italy's Eni group in gas projects. Ghana has most of its foreign-owned firms in Cluster 3. Most firms in extractive sectors, such as mining, do not employ large numbers of workers, tend to use less materials, and provide low wages. It may therefore not be accidental that such firms, even when they are entirely-foreign-owned, are associated with lower values along the performance dimensions analysed in this study. In southern Africa, the economy of Zambia presents another case analogous to Ghana. The most flourishing sector in Zambia is mining, which is mainly extractive, with the country's traditional exports of copper and cobalt accounting for over 70% of export earnings. Whereas many foreign-owned investments are found in this sector, irrespective of type, firms are likely to exhibit lower employment levels. The sector only employs 2% of the Zambian population. In 2014, in the Democratic Republic of Congo (DRC), the extractive sector only contributed 11% of employment but accounted for 95% of the country's exports (Transparence & Les, 2015). In Malawi, the sector contributed only 0.18% of employment in 2016 (Neumann & Kumwenda, 2017). The sub-section on case study firms in some of the countries in the data set lends credence to this claim, having cases in Ghana, Mozambique and Kenya, which has the largest number of firms in Cluster 4.

The relative differences between Clusters 2 and 3, (which confirm the hypothesised heterogeneity), together with the surprising result of firms in Cluster 2 (with less foreign ownership intensity) revealing superior performance outcomes relative to those in Cluster 3,

---

[30] Njikam (2018), Naughtin &Rankin (2016), and Rankin *et al.* (2005) have empirically confirmed that exporters are in most cases more productive than non-exporters.

are consistent across the three levels of this study's analysis. In the four-individual country-level analyses, diversity between clusters with foreign-owned firms is visible for all countries. In the appendices, results in Tables 3G1&3G2, 3H1&3H2, and 3I1&3I2 indicate that Clusters 2&3, 2&4, and 4&5 for Kenya, Nigeria, and Tanzania respectively exhibit firm-level heterogeneity among foreign-owned firms, as well as that clusters with less fdi-stake have higher values along those dimensions where differences exist between the clusters. At the third level of analysis, where the algorithm clusters separately firms from low-income and lower middle-income economies, diversity amongst foreign-owned firms is still consistently visible in both categories of economies. In Table 3J1 and Table 3J2 in the Appendix, Cluster 2 and Cluster 3 systematically differed. Also, Cluster 2, with lower fdi_stake, posted higher mean values on all the performance dimensions, further attesting to the sort of performance contradiction noted earlier. A similar picture was mirrored by results in Table 3K1 and Table 3K2 for lower middle-income economies for about half of the variables under consideration.

The consistency in heterogeneity and the surprising result is reflective of findings from other empirical studies that apply different methodologies in both SSA and elsewhere. In relation to heterogeneity, utilizing the World Bank Enterprise Survey data, Pfeiffer *et al.* (2014) also finds firm-level heterogeneity in terms of horizontal productivity effects amongst emerging and traditional foreign direct investments in SSA. Elsewhere, Coniglio *et al.* (2015) estimate a CES production function using MM and OLS methods. In their analysis of wages and employment, they find significant dissimilarities between foreign-owned investments from developed economies (North) and those from the South (developing economies) in skill intensity and wages[31]. Coniglio *et al.* (2015) also find Chinese-owned firms employing more workers (indicator of firm size) than foreign-owned firms from other countries, a further confirmation of diversity within foreign-owned investments. In the same vein, in a study of Greek firms by Georgopoulos & Preusse (2009), econometric results indicate that acquisitions exhibit specific signs of excellence over greenfield investments in terms of market share, firm size, capital intensity and product differentiation. This alternative scholarly evidence supports this study's clustering results regarding the significant differences existing between foreign-owned investments, the sort of performance contradiction that was found pervading foreign-owned investments; and it answers the second objective of this chapter positively i.e. whether

[31] See also studies by Ghosh and Roy (2018) and Nishiyama (2010).

heterogeneity does exist at firm level as previous study findings established, and whether it is more likely among foreign-owned investments, domestically-owned firms or both.

In Table 3B in the Appendix, considering that Cluster 1 is composed almost entirely of domestically-owned firms, it is noticeable that, out of 435 foreign-owned firms in the lower middle countries, 61% were in Cluster 3 and 13% in Cluster 2. It is also visible that, out of the 401 foreign-owned firms in Uganda, 7.2% were in Cluster 2, and 84.8% in Cluster 3. Lastly, it can be noticed that out of about 1475 foreign-owned firms in low-income countries, excluding Uganda, 15.4% were in Cluster 2, 64% were in Cluster 3, 20% were in Cluster 4 and only 0.1% were in Cluster 5. Cluster 3 has the largest number of foreign-owned firms. Uganda and the other two economy groupings, being low-income and lower middle-income economies, had respectively the largest percentage of foreign-owned firms in this cluster. This implies that foreign-owned investments in Uganda and elsewhere in SSA tend to exhibit homogeneous characteristics as reflected along the variables under consideration. From the same perspective, it is clear that domestically-owned firms are featured by similar firm-level characteristics in Uganda and elsewhere in SSA. If this were not the case, the Ugandan firms or other country-specific firms would have been clustered differently. This supports the assertion that, irrespective of whether the algorithm is applied to firms from Uganda, to low-income economies, or to lower middle-income economies, intra-foreign-owned firms' differences are likely to be consistently exhibited by firms owned by foreign investors; and they are more likely to be a cross-cutting firm-level feature across SSA and probably developing economies. In summary, the result is not peculiar to any country, including Uganda. This result supports, from the perspective of foreign direct investments, this study's earlier hypothesis that Ugandan firms may not be unique, and that the earlier results of Chapter 2 are likely to hold elsewhere in relatively similar economies.

Cluster 4 (6.8%) - moderately foreign, large size/capital intensive firms

With reference to Table 3.4, It is evident that firms in Cluster 4 are also foreign-owned but with a mean fdi-stake of 41.7%. These firms are equally likely to be export oriented on average and more likely to be M&As in terms of foreign-ownership categorization if this study's earlier assumption was to be invoked. Although by this study's operational definition, firms in this cluster qualify to be categorised as foreign-owned, they are inherently more domestic in nature than foreign-owned firms in other clusters. Despite this domestic leaning feature, firms in Cluster 4 post mean real values, which are above the SSA averages for all the variables considered in the analysis. This is a pointer to the significance of foreign ownership stakes at firm level, irrespective of the intensity of such a stake. Table 3.4 illustrates that more foreign-owned firms' heterogeneity is prevalent between Clusters 3 and 4, relative to what we see between Clusters 2 and 3. These two clusters are significantly different from each other on all variables considered in the analysis. Once again, except on the initial investment variable, Cluster 4 with less fdi-stake still posts higher mean values than Cluster 3, which is more likely to consist of firms that are completely foreign and hence likely to be greenfield investments. This result echoes the surprising result discussed earlier, although some available literature[32] tends to support the hypothesis that more foreign-ownership intensity is associated with higher levels of firm performance. However, I attribute this sort of performance contradiction to probable cherry-picking tendencies as earlier discussed.

Employee training, a proxy for labour skill intensity, had results for firms in Cluster 4 and Cluster 3 partly mirroring the findings of Coniglio *et al* (2015). These scholars, estimate a CES production function using both MM and OLS and find that the presence of a local partnership in a foreign-owned firm is associated with a more-skilled worker force (Coniglio *et al.*, 2015: 1256). Firms in Cluster 4 have more local partnerships[33] than those in Cluster 3. These firms are therefore likely to offer employee training, something that ultimately enhances the firm's skill intensity for the workforce. Moreover, formal employee training in foreign-owned investments in SSA is still low just as in local firms (Chen *et al.*, 2015). These results are demonstrated in Table 3.2 Clusters 1, 2, 3 and 4; and in the p-values in Table 3C in the Appendix, confirming the statistical significance of the differences. The correlation matrix in Table 3.2 also indicates zero correlation between ownership and training.

Overall, taking all clusters together, it is noted that firms in Clusters 2, 3 and 4, being the clusters with foreign-owned firms, are more likely to be export-oriented, capital-intensive and have higher levels of labour productivity. This trinity of firm-level characteristics among foreign-owned investments in host economies has been confirmed by previous empirical findings in studies such as Wagner (2005), Rankin *et al.* (2005), Greenaway and Kneller (2007), Bouras & Raggad (2015) and Christos *et al.* (2016). I mostly found similar evidence in Chapter 2, implying that the findings of this chapter support the hypothesis regarding the

[32] See studies by Javorcik (2013) and Garavito *et al.* (2014).
[33] By the mere fact that they are not fully foreign-owned

absence of peculiarities between Uganda and the SSA case and, more generally, developing economies.

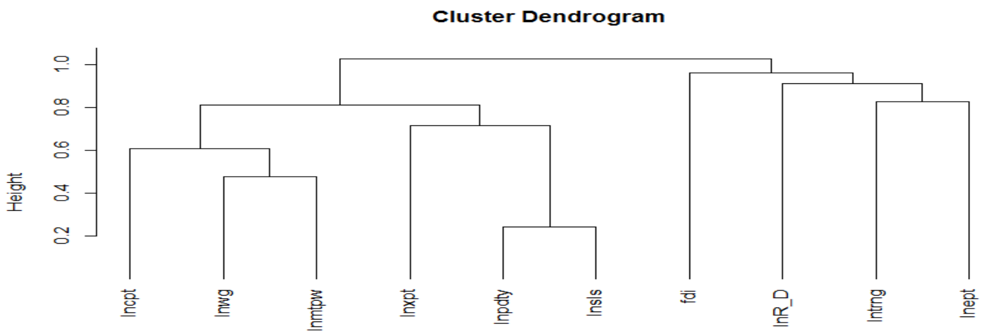
3.3.4 Further validation checks on the results

Apart from using Gower's distance measure to mitigate the effect of extreme variable collinearity on this study's results, I performed a clustering of variables process. This yielded principal clusters of variables which I used to test whether similar cluster solutions could be achieved. I investigated this by using the silhouette plot based on the new synthetic variables. Using the ClustOfvar techniques, I generated four clusters of synthetic variables. The first cluster consisted of six variables, being capital intensity, wage, material per worker, exports, labour productivity and sales. The second and third cluster respectively consisted of fdi stake and research investment, while the fourth cluster consisted of employment and training.

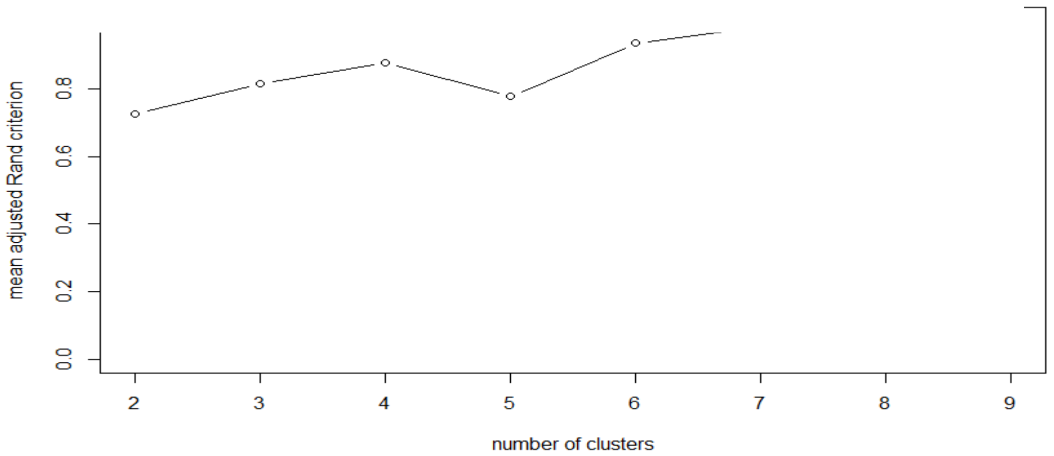
In Figure 3.7, graphical illustrations of the cluster dendrogram for the variables (Panel A), the stability plot for the partitions (Panel B) and the silhouette plot (Panel C) for the firms clustered based on the synthetic variables, are shown. In Panel A, from left to the right, abbreviated variables are capital intensity, wage, material per worker, exports, labour productivity, sales, fdi stake, research investment, training and employment. Sales are more highly correlated with productivity (see Table 3.2) and are equally linked together in Panel A. Wages are more linked to labour productivity, but in Panel A are paired with material per worker after productivity is clustered with sales. There is a close link between the cluster dendrogram and the correlation matrix in Table 3.2. Looking at Panel A, many clusters of variables can be seen; although *fdi*, in particular, constitutes its own cluster once the cluster numbers exceed two. In Panel B, one sees the stability plot, with the first highest peak at 4 clusters, just before the break at 5 clusters. The higher the peak, the better the structure; however, in this study's case, I ignore any structures after the break as any cluster number beyond five yields more single variable clusters, whereas my aim was to reduce redundancy.

I utilised the four synthetic variables generated by the variable clustering process to cluster the SSA firms again, using the Euclidean distance measure but still linking the firms using Ward's method. In Figure 3.7 Panel C, the silhouette plot of the resulting clustering process is shown. The ASW in this case is far lower than 0.5; however, it should be recalled that, in this case, ASW was measured based on log-transformed variables. The highest ASW was attainable with two clusters at slightly above 0.3. A 5-cluster solution, which mirrors this study's earlier results is attainable at an ASW slightly lower than 0.3, which indicates a superior cluster solution when compared to earlier results. This superiority is because the ASW in this case is based on logged hybrid variables. This confirms the validity of the clustering results earlier discussed.

Panel A



Panel B



Panel C

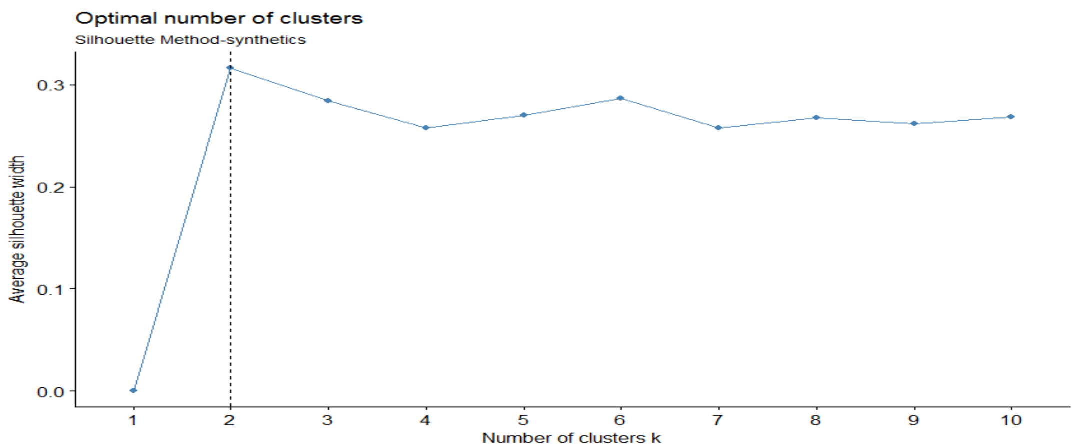


Figure 3. 7: Variables cluster dendrogram, stability plot and silhouette plot

3.3.5 Conclusions from the Clustering Results

In the previous sub-section, I conducted a cluster analysis of firms in SSA. The analysis yielded results similar to findings in Chapter 2 but on a much broader data set and for a multi-country scope. The analysis revealed not only systematic differences between foreign-owned and domestically-owned firms but also intra-foreign-owned firms' heterogeneity. This empirical finding is common in earlier studies on foreign investments which utilised different methods of analysis and data sets, often across geographic regions. The element of firm-level heterogeneity between foreign-owned firms was not an explicit result in Chapter 2, probably due to the size of the data set. This heterogeneity is more likely between groups of foreign-owned firms that exhibit significant variations in the intensity of foreign ownership and exportation. It could be that these two variables are the leading influencers in the segmentation process of firms in the data.

Most importantly, the results in the previous sub-section confirm that such systematic differences between foreign-owned and domestically-owned firms are unlikely to be unique to a country case (for instance, Uganda), but rather a firm-level feature that is likely to characterise foreign-owned and domestically-owned enterprises elsewhere. Additionally, results also reveal that firm-level heterogeneity is a feature that characterises foreign-owned investments. As other empirical studies have confirmed, the likelihood of firms with less than 100% foreign ownership performing better than their counterparts who are completely foreign is high. This is consistent at the three levels of cluster analysis. A key explanation for the rather contradictory performance differences between partly foreign and fully foreign owned firms as described in this chapter is the probable cherry-picking pattern but other explanations are available in the literature.

The percentage difference in sector composition is on average higher in clusters with foreign-owned firms, with the manufacturing sector contributing a larger percentage than the service sector. Cluster 1, with 99.95% domestic firms, has the lowest percentage difference of 11.6%; and Cluster 4 has the highest percentage difference at 33.4%. This suggests that foreign-owned firms are more likely to be found in the manufacturing sector than the service sector, and that there are fewer domestic firms in manufacturing in SSA. This can be attributed to a myriad of factors, with technological inadequacy, human capital and resource constraints contributing to this. This further reaffirms the hypothetical belief regarding the ability of foreign direct investments to contribute to the much-needed structural transformation in developing economies through the creation of employment opportunities in modern sectors. This is not the case in Rwanda where in 2013, foreign direct investments were 41% in the ICT sector, 20% in the financial and 19% in manufacturing (Chen *et al*., 2015).

Most importantly also, the clustering algorithm consistently yielded results with similar evidence at three different levels: at SSA-level, for certain specific-country clustering, and for separate clustering based on low-income and lower middle-income economies. This shows that this study's results are robust. The composition of various clusters, both at the first and at the third level of analysis, attested to the fact that firms in Uganda were not unique to other firms in relatively similar economies. The algorithm yielded clusters that indicated homogeneity between firms, irrespective of country of location or of whether the country was a low-income economy or lower middle-income economy. Therefore, findings in Chapter 2 regarding the salient differences between foreign and domestically-owned firms were not specific to Uganda but applicable to similar economies elsewhere.

The clustering results, however, did not show anything regarding the relative importance (segmentation power) of foreign ownership stake and other variables in firms' segmentation, or about the consequent formation of clusters. Findings on variable importance in segmenting firms form part of the results in Chapter 2, where a country-specific analysis was undertaken. I therefore investigated this element of variable importance to ensure consistency in fulfilling Chapter 3's cardinal objective. In this endeavour, I employed a CART analysis technique.

3.3.6 Classification and Regression Tree Analysis (CART) of SSA firms' segmentation

In the previous sub-section, results suggested that, along the selected key performance and characteristic variables, the agglomerative hierarchical clustering algorithm segments firms into two clusters at the penultimate stage. It is shown in the cluster dendrogram and the silhouette plot, in Figure 3.5 and Figure 3.6 respectively. Additionally, a close inspection of the dendrogram showed that before the penultimate segmentation, multiple clustering at the third and other preceding stages occurred, appearing visually more concentrated on the right than on the left. The penultimate segmentation probably yielded the predominantly domestic firms' cluster, while foreign-owned firms were further grouped into different clusters in the rest of the stages.

Understanding further the variables that hugely informed this clustering process and the consequent cluster formation was part of my keen study interest. This would shed light on whether ownership status (fdi-stake) plays a significant role in firm classification. Understanding the extent to which ownership status (fdi-stake) influences firm segmentation provides preliminary explanation as to whether or not, the observed systematic differences are significantly associated with foreign ownership itself, which exact extent and effect I later investigate in chapter 4 of this thesis. This would further illustrate whether the findings of Chapter 2, which attached relative importance to foreign ownership and other variables, are robust for the SSA case. This would support my earlier assertion that findings in Chapter 2 were not peculiar to Uganda but applicable elsewhere by providing additional supportive evidence to this chapter's first cardinal objective. Succinctly, I tried to understand which variables had the largest influence on how the algorithm grouped the firms in the data set. The visual inspection of the dendrogram only enabled me to speculate on how the clusters were formed, not the key variables that informed the process or the influence of each variable. It is for this reason that I executed a CART analysis on the SSA data. Classification and Regression Tree Analysis, CART, is a simple yet powerful analytic tool that helps determine the most "important" (based on explanatory power) variables in a particular dataset, and can aid researchers craft an effective explanatory model. Using a splitting rule based on a Gini Index, the algorithm solved the maximization problem specified in equation 2.11. While CART is infrequently applied in the economics field, this analytical tool and its conceptual fabric is firmly entrenched in health research, for the most part in epidemiological and clinical settings.

I developed regression trees and the most important variables (variables that hugely influence the clustering at various stages) were identified. I used *rpart*, an R package, which generated comprehensive results of CART, including variable importance (VIMP) based on percentage contribution, or explanatory power, in the segmentation of firms. VIMP of variable X_v is a measure of the increase, or reduction, in prediction, or misclassification, errors on the data set if X were not available. A large VIMP indicates a variable with classification or predictive ability, while a zero or negative VIMP shows non-predictive variables (Ishwaran *et al*., 2008:11). Using the *rpart.plot* package, I produced an illustrative CART tree for visual inspection of how firms were segmented. The CART regression model is specified as:

as.factor (groups) ~ employment + lbr_prdvty + fdi_stake + capital_intensity + material_perworker +sales+ wage + employee_training + research_investment + exports

In this formulation, the dependent variable groups is the cluster, a categorical variable taken by the R package as a factor. In Table 3.5, the relative variable importance in segmenting firms is shown.

Table 3. 5: Variable importance in segmentation of firms in SSA

Variables	*Absolute contribution*	*Percentage*
fdi_stake	6983.02	92.0
Exports	358.41	5.0
employment	158.87	2.0
Sales	67.14	1.0
material per worker	29.67	0.0
employee training	16.18	0.0
labour productivity	11.62	0.0
research investment	2.6	0.0

Source: Author's own results based on AIS Survey Data, 2010

With reference to Table 3.5, it is noticeable that fdi_stake had the highest absolute level of importance in segmenting SSA firms. This variable accounted for over 90% of the segmentation process of firms and for the resulting formation of firm clusters. Exportation, employment and sales followed closely, respectively contributing 5%, 2% and 1% to the segmentation process of firms in the data set. The rest of the variables, although with visible absolute contribution, are associated with 0.0% contribution to the clustering process. In Figure 3.7, a pruned tree-like visual illustration of the CART results is shown.

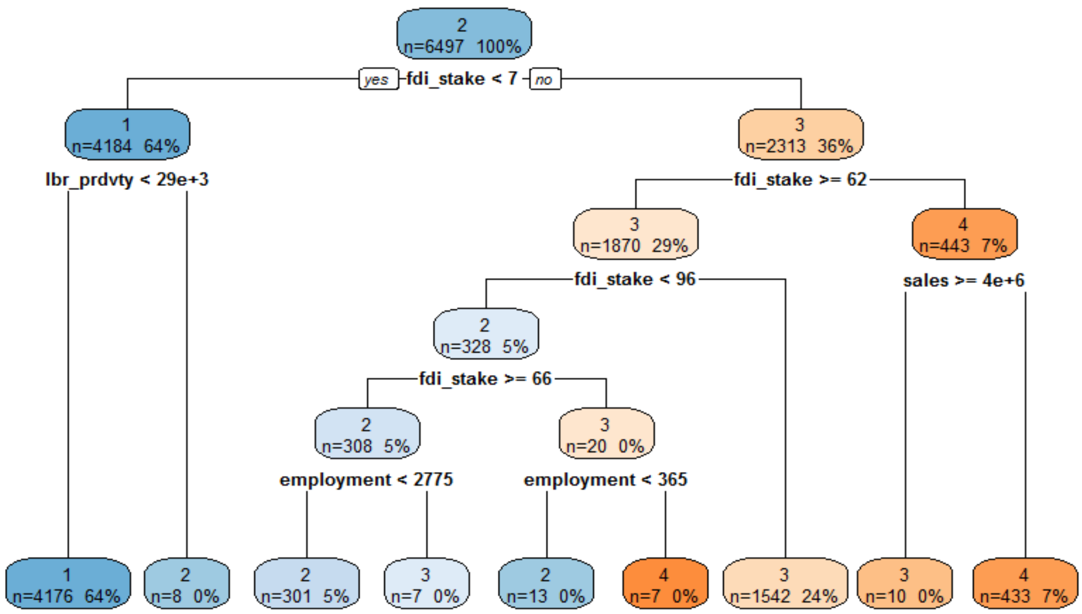


Figure 3. 8: A CART tree diagram based on SSA dataset

In Figure 3.8, in the top box, we see 100% composition of the data set and the total number of firms (n = 6497) which are segmented based on the variables selected for analysis. The "yes" and "no" options correspond to the left and right branches, respectively, at each segmentation stage. By reading from the top[34], we notice that at the first, top level of segmentation, firms are segmented based on their foreign ownership intensity (fdi_stake variable). At this stage, 4184 or 64% of the firms were clustered away on the left. These firms were very homogenous along the majority of variables but were hugely similar because their fdi_stake was less than 7%. The two foreign-owned firms clustered in this group are more likely similar to domestically-owned firms along the rest of the variables than other foreign-owned firms are. The mean fdi-stake of these two firms is 15%, below the SSA mean value of 30.6%, only one of them exports, neither invested in employee training or R&D, and they were also less capital intensive; all of which are features of most domestic firms.

It is shown that after the top-most segmentation, Cluster 1 was further segmented, based on labour productivity. However, even this second segmentation is negligible when considering the fact that the variable, which accounted for the split at the node, was rated at a VIMP of 0.0%. Indeed, after this split, only 8 firms (0%) are clustered away to the right. This is an indicator of their (domestically-owned firms) homogeneity at that stage relative to the foreign-owned firms on the right-hand-side (RHS) of the tree. Segmentation of firms on the RHS at the second stage was still based on foreign ownership intensity, with those having a minimum of 62% fdi-stake segmented to the left and those below 62% clustered to the right. Firms with fdi-stake of at least 62% but less than 96% were clustered to the left at the next stage while the totally foreign, likely Greenfields investments with fdi-stake more than 96%, were clustered away to the right forming a cluster of 1542 (24%) firms. Those firms with less than 62% fdi-stake were later segmented based on lower foreign ownership intensity, employment, exportation potential, sales and other variables to yield further clusters. Some of these variables may not appear on the CART tree diagram due to pruning to reduce the size of the tree.

Significant to note is that exportation is not an important classification variable between foreign and domestically-owned firms, but it is within foreign-owned firms. This finding may be attributed to the fact that some foreign investors target markets in host economies, making them

[34] Although the clustering algorithm was hierarchically agglomerative, the tree is easily read from the top.

foreign domestic-oriented firms, while others target regional markets as well as the local markets in host economies. Available empirical literature on foreign direct investments suggests that most foreign direct investments in SSA, especially in manufacturing, are majorly market-seeking (Chen *et al.*, 2015; Henley, Kratzsch, Külür & Tandogan, 2008: 13).

In this CART analysis sub-section, more insights are shown regarding how the clustering process segmented firms. Above all, I established the relative importance of variables in the process of segmenting firms in SSA. I therefore provided more evidence regarding the hypothesis that firm ownership status (being foreign or local) has the greatest separation role in firm clustering. I also showed that exportation is more of a separating variable between foreign-owned firms themselves than between domestically-owned firms and foreign-owned firms. These results echo the finding of Chapter 2 where a smaller country-specific data set was analysed.

The quantitative analyses in the previous two sections (cluster and CART analyses) provide further evidence on the factors that are likely to be associated with foreign investments, as well as on how these factors systematically distinguish foreign-owned firms from domestically-owned firms. Specifically, findings so far reveal that even in the SSA case, observed performance differences between foreign and domestic firms are likely to be due to ownership status and not mere subgroups/sectors where these firms are anchored. These findings tend to imply that variations in firm-level performance are likely to be associated, to a recognizable extent, with the structure of firm ownership. Ownership is more of a structural characteristic variable and findings so far may not tell whether the observed differences are solely (or to which extent) attributable to ownership status. As opposed to the current multivariate exploratory investigation, this calls for causal analysis, a key inquiry endeavoured in chapter 4 of this thesis. Also, noticeable so far is that some implications derived from findings of cluster analysis, especially on intra-firm differences among foreign-owned firms, are premised on how foreign ownership might have resulted. This context, other than being assumed based on the level of foreign-ownership stake on average in a particular cluster, is not reliably confirmed by my analysis. By implication then, unless otherwise, results and associated conclusions especially on firm-level heterogeneities, are unnecessarily subject to circumspect interpretation and application. On these grounds, I attempt to add credence to this chapter's findings and inherent implications by undertaking a qualitative case study analysis of carefully selected foreign-owned firms in SSA. These cases, selected from countries that form part of the data set used, provide real life context under which foreign ownership usually arise among firms in

SSA, how performance usually evolves over time after foreign investors gaining stakes in local firms or after setting up new plants in case of Greenfield investments. The thought was that having contextualized how ownership does come into existence, how performance of such firms has evolved overtime, and possible challenges faced, the observed varying average performance and effects of firms will be easier to envisage. This lends more credence to my findings and conclusions. In the next section, this analysis follows;

3.4 Case Study Analyses

In this analysis, I selected case studies representative of all key sub-regions of SSA. Cases of firms that were originally domestically owned but later acquired by foreigners were historically traced. The aim was to answer questions related to; how foreign ownership emerged and what kind (merger, acquisition, or greenfield), circumstances that led to kind of ownership, and how performance evolved especially after; acquisition, merging or first-time establishment of the firms in question? In a further step, these firms were integrated in my clustering algorithm to confirm whether they would be clustered in the expected groups. Apart from Lafarge Cement, all the analysed firms were grouped as expected. Following is a case-by-case analysis.

3.4.1 Ghana Rubber Estates Limited (GREL)

GREL is a natural rubber-producing firm in Ghana. It is 60% foreign owned and 40% locally owned. GREL's ownership mix qualifies it as a foreign-owned firm but with partial foreign ownership. This enterprise is therefore similar to the majority of firms analysed in the cluster analysis section. The similarity was further confirmed when, using firm-specific data for GREL[35] in 2010, I ran the algorithm again, resulting in the firm being grouped in the fourth cluster. In Ghana, the firm accounts for close to 95% of the total natural rubber produced (Sutton & Kpentey, 2012). GREL's position in the sub-sector and its similarity to firms studied in cluster analysis make it a suitable case study to examine. Moreover, GREL's output is largely exported both within and outside Africa, giving the firm a good feature (export status) that is associated with most foreign-owned firms in the literature.

The current GREL was a culmination of government's nationalisation policy and consequent take-over of a small (923 hectares) private plantation established by R.T Briscoe. Briscoe was a Ghanaian entrepreneur. The private plantation was established in 1957 at Dixcove. Under the

[35] Data obtained from annual integrated reports at www.siph.com

nationalization policy, the private plantation was transformed first into the Agricultural Development Corporation (ADC) and later (in 1962) into a State Firms Corporation (SFC). This process turned the formerly private firm into a purely state-owned enterprise. However, financial constraints, emerged, forcing the government of Ghana to look for a capable financier. Firestone Tyre Company of the USA took on this role. This led in 1967 to the formation of GREL as a joint venture firm owned by the state and Firestone.

In 1990, Firestone sold its shares to the government (Sutton & Kpentey, 2012). This transaction once again transformed the enterprise into a purely state-owned enterprise. During this period, the enterprise entered a period of drastic decline, and plantations were almost abandoned. The Ghanaian state, however, agreed to a financing agreement with Agence Francaise de Development. The agreement ushered in proper management and culminated into the rehabilitation of the rubber plantations. The agency was also meant to construct a new rubber processing plant at Apimenim[36]. However, starting from the late 1980s, Ghana embarked on a divesture program for some of the state-owned enterprises. Consequently, in 1997, after GREL had been rehabilitated, it was privatised, with majority of the stake (60%) sold to Société Internationale de Plantations d'Hévéas (SIPH), a French-based company. The state retained 25% and a Ghanaian firm, Newgen Company Limited, took over the remaining 15%.

The government's action prior to 1967 and its policy shift prior to 1997 opened the way for participation in the ownership and management of the enterprise by foreign investors. In the first phase, Firestone gained ownership of part of the enterprise, while in the latter phase, majority ownership was sold to SIPH. In both cases, foreign ownership was associated with performance benefits for the enterprise. During Firestone's ownership reign, GREL inherited about 39 000 hectares of plantation. However, from 1988, planting of additional hectares of rubber was embarked upon to boost output. And although Firestone had left by 1996, 4 000 additional hectares of rubber plantation were in place (Sutton & Kpentey, 2012). Foreign acquisition of the enterprise by SIPH further improved its performance. Through collaboration with Michelin, SIPH accessed vital technical assistance for the firm. Additionally, this collaboration expanded the market for the firm's output. Michelin is the largest purchaser of the enterprise's output since rubber is a key input in Michelin's tyre manufacturing activities. Improvement in the enterprise's performance was also exhibited in terms of employment,

[36] www.grelgh.com, 2013 copyright

revenue and output. According to available literature, this improvement, which has been manifest ever since partial acquisition by SIPH, continued over recent years. For instance, in terms of output growth, GREL's own and purchased rubber production rose from below 15 000 tons in 1997 to over 50 000 tons by the first quarter of 2017 (Sicom, 2017a,b). This is relatedly confirmed by the enterprise's growth in revenue from below US$ 10 million in late 1990s to annual turnover of US$ 41.2 million in 2010 according to (Sutton & Kpentey, 2012). By the first quarter of 2017, the firm's sales reached 54 000 tons, yielding a turnover of €94.7 million (Sicom, 2017a).

In an attempt to probe what has accounted for GREL's performance since the takeover of its partial stake by SIPH, several factors can be singled out. These factors can be categorised as firm-specific factors, factors related to domestic economic policy in Ghana and natural factors, given that GREL is anchored in the agro-industrial subsector. However, firm-specific factors are most pertinent to our study.

In terms of firm-specific factors, the majority owner of GREL, SIPH, is one of the largest firms in production and marketing of natural rubber globally, and is a leading firm in Africa. SIPH's global size and leadership in Africa is manifested in its turnover and international presence. For instance, SIPH group had a turnover of €422.3 million in 2011 (Vignes, 2012: 11). Besides the group's business in France and Ghana, SIPH has other affiliates in other African countries, namely Cote d'Ivoire, Liberia and Nigeria. This gives the group reliable economic potency to leverage in order to ensure smooth operations in its affiliates like GREL. Foreign-owned firms have been known to be technologically superior compared to domestically-owned firms in host economies. Acquisition by SIPH provided GREL with the vital technical expertise that forms the foundation of any manufacturing firm in any sector. Currently, Michelin provides this technical prowess.

SIPH sells most of its output internationally, with Michelin as the largest buyer. This characteristic accords GREL export status, a position that typifies many foreign-owned investments. As a largely export-intensive firm, GREL's performance can thus be linked to its export status. Exportation is aided even further by the nature of GREL's product. Rubber is one of the most commonly used goods in almost every industry. Its use ranges from tyre manufacturing to aviation, education, sports, health and engineering, all of which ensures ever-rising demand globally. For instance, global consumption of natural rubber increased by 5.2% in the first seven months of 2018 alone, according to the Association of Natural Rubber

Producing Countries (ANRPC) report for 2018. Rising global demand is a factor that favours export-oriented firms like GREL that serve global markets. Numerous empirical findings in studies such as (Njikam, 2018: 7) confirm that exporting firms post higher levels of performance. Some of these empirical studies[37] have specifically been conducted on Ghanaian firms. Cluster results in the previous section also tend to suggest the same.

Besides firm-level characteristics, the probable role of natural climatic conditions and physical resources cannot be ignored regarding the enterprise's performance. This is because the key input in the firm's production processes is agricultural in nature, and climatic conditions and other related natural events therefore play a key role. Barthel *et al.* (2008) analyse the characteristics and determinants of foreign direct investments in Ghana. Among their empirical findings is the significance of natural and physical resources in attracting foreign investment in Ghana. From the policy perspective, liberal market policies by Ghana coupled with political stability over the years has enabled private sector growth and first-rate performance of firms like GREL.

3.4.2 Lafarge Cément Zambia (LCZ) Plc

LCZ is a cement-producing firm in Zambia. It is 84.5% owned by LafargeHolcim group of France. Its partial foreign ownership status is a reflection of many of the firms in the data set used in cluster analysis for this dissertation. Additionally, LCZ is anchored in one of the major sectors of the Zambian economy, being the mining and extractive sector. Relative to GREL's rubber, the product produced by LCZ cannot easily be exported given its bulk nature. This makes LCZ an instance of a partially-foreign-owned firm that is likely to be less export-intensive but that probably has other features of foreign investments common in the literature. We use actual data on selected variables for this firm[38] and feed them in the algorithm. The algorithm clusters this firm away from the rest, forming a single-element cluster. This probably is indicative of this firm's uniqueness. LCZ's ownership mix, sector type and likely export status are some of the elements that make it an appropriate case to examine in the context of this study.

The current LCZ was originally a state-owned firm called Chilanga Cement established in 1949 (Sutton and Langmead, 2013). The chief aim of establishing Chilanga Cement Factory was to

[37] See Abor (2011): 'Does export status and export intensity increase firm performance?'
[38] I obtained these data from the Lafarge Group archives for 2010.

supply cement for the construction of the Kariba Dam wall. Cement production commenced in 1951.After nearly two decades, a second plant was constructed at Ndola in 1969. By 1970, deficiencies in the country's strategy of state ownership of enterprises and direct involvement in economic activities had begun to emerge. These deficiencies were worsened by the 1973 oil crisis. This led to a re-thinking of the state-ownership strategy, paving way for denationalisation of state-owned enterprises. As a result, Chilaga Cement firm was denationalised in 1994, with the majority stake acquired by the Commonwealth Development Corporation (CDC). This policy action ushered in the first foreign ownership entity in the Chilanga's ownership mix. CDC formed the Pan African Cement (PAC) in early 2001, which had ownership stakes of 50.1% in Chilanga Cement, 75.2% of Portland cement in Malawi and 58% of Tanzania's Mbeya Cement respectively (Sutton and Langmead, 2013). CDC acquired the above ownership stakes as part of the divesture programs for state-owned enterprises of each respective governments. In 2001, Lafarge Group acquired PAC from CDC, and in 2007 changed the name of the Zambian firm (Chilanga) to LCZ. Lafarge Group currently owns 84.5% stake in the firm.

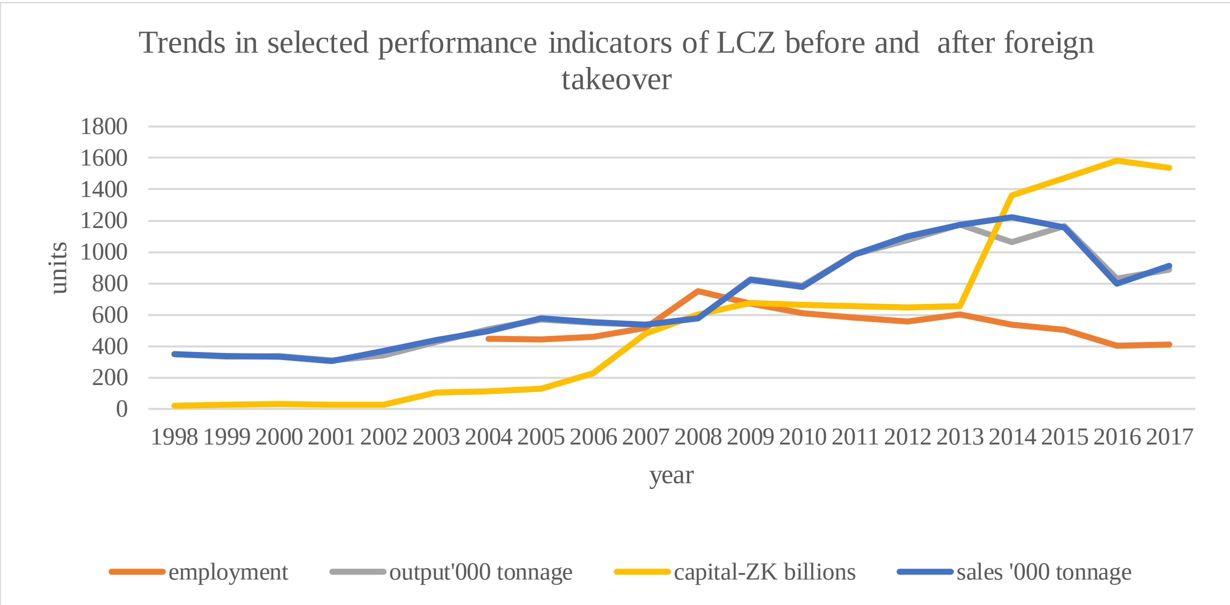
Foreign acquisition of Chilanga Cement was and continues to be associated with improvements in the enterprise's productive capacity and general performance. This was very evident with the takeover by Lafarge. Improvements in the enterprise's performance manifested in areas of product development, employment growth, enhancement of production capacity, increased sales (and hence turnover), and corporate social responsibility (CSR). In 2003, LCZ launched the firm's flagship general-purpose cement brand, Mphamvu. This was followed by establishment of another plant, Chilanga2 Factory, at an investment outlay of US$ 120 million in 2008 in order to meet growing domestic demand for cement. As a result, the firm's productive capacity increased to a magnitude of 1 300 000 metric tons of cement per year (Sutton and Langmead, 2013).

Between 2010 and 2014, LCZ undertook further product development, launching over three new products. These included: Powerplus to cater for heavy construction needs; Supaset, a fast-setting cement to cater for concrete product industrialists and block architects; RoadCem for road construction; and WallCrete for masonry uses[39]. These various products underscored LCZ's ability to meet varying consumer needs in the cement industry; and attested to

[39] https://www.lafarge.co.zm/1_7_6-History

productivity changes experienced by the enterprise after foreign acquisition by Lafarge. Moreover, in the same period, a new aggregates plant was commissioned at Mapepe, further enhancing the productive capacity of the enterprise.

Reflecting the findings of most studies on foreign-owned investments, LCZ prioritised research and development expenditure. This probably explains the firm's success in product development highlighted above. In 2014, LCZ commissioned its first concrete products laboratory in Zambia shortly before launching PowerCrete cement, specifically for use in the mining industry. In the area of CSR, in 2015, the firm established Lafarge Foundation Zambia, and Lafarge Zambia Plc officially became a member of LafargeHolcim Group. These milestones in the life of the enterprise attest to the effect that foreign acquisition probably had on its performance. Figure LCZ1 shows trends in selected performance indicators of LCZ. We illustrate trends in five performance indicators three years before Lafarge's acquisition and sixteen years after the acquisition. In the figure, employment is measured in hundreds of workers, output and sales of cement are in thousands of tons, capital investment captured by the value of equipment, plant and machinery is measured in billions of Zambian Kwacha.



Source: Author's own illustration based on Lafarge Cement Zambia Plc annual reports

Figure 3. 9: Trends in some performance variables for LCZ, 1998 - 2017

In Figure 3.9, it is evident that for at least three years prior to Lafarge's acquisition of the enterprise, sales and output were on a declining trend and capital levels were nearly constant. After 2001, we notice a rise in the selected indicators, with capital investment picking up late in 2003, probably because of the fact that its effect takes time to be manifested. Even with missing data, employment shows an increasing trend from 2004 but begins to decline in 2008 until it stabilises in 2016. However, capital continues to rise over the years as output increases too. This is suggestive of the firm's inclination to capital-intensive production technology, a key feature of foreign-owned firms revealed in cluster analysis and most empirical studies. In summary, it is clear that the enterprise witnessed a period of improvement after acquisition by Lafarge.

The case of LCZ and its stellar performance can be associated with both firm-specific features and enabling features of the host economy. In the former, pertinent features are mainly those of the acquiring foreign firm, Lafarge. In terms of firm-specific features, Lafarge Group is the world's largest manufacturer of cement and roofings, second largest producer of aggregates and concrete products, and third largest in global production of gypsum. Firm size has empirically been proven to be linked to performance. In 2002, Lafarge Group's sales amounted to €14.6 billion and the Group has presence in over 70 countries globally (Geho, 2018).

Given its global position, the group has vast experience in the cement industry. Lafarge's takeover of Pan African cement (PAC) from the Commonwealth development corporation (CDC) represented the Group's vast expansion in the developing world, and SSA in particular in recent years. This expansion can be attributed to the Group's intent to fulfil its broader strategy, which is both informed by the Group's economic forecasts of the SSA's market and pivoted on the following: '…a focus on countries with empowered local markets, which are profit and loss accountable by 2022'.[40]

The statement below indicates the experience of the acquiring firm in the industry, its management capacity in terms of strategy formulation and forecast, besides its economic potency:

> *"The four countries that constituted PAC, had by 2000, a combined cement consumption of 2.5 million tonnes, which had experienced significant growth in the last five years and was forecasted to develop at an average of 4% over the next five years".* (Lafarge press release, 2000)

Economic potency does not only allow foreign-owned firms such as Lafarge to undertake required investments but also positions such firms better as far as access to investment finance

[40] LafargeHolcim, 2017 annual report, page 8

from lending institutions and at favourable rates is concerned. Domestically-owned firms are usually financially constrained. This easy and favourable access to finance is equally echoed by Aliber's theory in this thesis' subsection of theoretical background in chapter 2.

Evidence in literature shows that Chilanga was initially performing averagely well as indicated by that successive addition of kilns in 1956, 1957 and 1969. Additionally, a new plant was constructed at Ndola with another kiln added in 1974 (Sutton and Langmead, 2013). Despite these milestones, several setbacks later undermined its performance and disrupted sustainability. Such setbacks included management and financial flaws associated with state ownership of parastatals. These were coupled with the oil crisis noted earlier. In Figure LCZ1, the variables indicate diminishing trends three years before 2001. Such flaws and shocks could easily be side-stepped through foreign investment firms like Lafarge, given their superiority in terms of finance, management and technology.

Also, firm-specific to Lafarge is the capital-intensity attribute that is vital to any enterprise's performance in a sector that is more extractive. Cement production is capital-intensive; and in this partly-foreign owned venture, the acquiring firm, Lafarge Holcim Group, is a global leader in the cement industry with the latest technology and plenty of capital equipment and investment funding. In Figure LCZ1, we notice that after Lafarge's take-over, the value of capital outlay takes on an increasing trend, indicative of improved investment in the enterprise. As expected, similar movement in output and sales accompanies this trend.

Away from firm-specific enablers of LCZ's performance, features of the host economy are worth mentioning. Growth in domestic demand for cement from both the government and the private sector make Zambia an attractive destination for foreign direct investments targeting such a sub-sector. This is more so for firms focused mainly on meeting the commonplace domestic demand for cement. Regarding government demand, even the most recent new entrants in Zambia's cement market, notably, Central African Cement Ltd, Dangote Cement Industries Ltd, Sinoma Mpande Limestone Ltd and WEYE Construction Materials, among others, have hugely been enticed by both the pending and ongoing infrastructural developments by the Zambian government. Indeed, available evidence from literature has it that:

> *"The recent expansion of Zambia's cement industry is partly fuelled by expectations of an infrastructure boom. In order to diversify the economy, government is keen on developing key infrastructure such as the Kazungula Bridge that will connect Zambia to Botswana to enhance intra-regional trade. Zambia is*

also in the middle of a national irrigation project to establish three hydroelectric dams at Mwomboshi, Chisamba district, Lusitu, in the south, and at Kafulafuta in the Copperbelt" CemNet.com, (Insights, 2018)

Besides government, private sector demand is also worth mentioning. Private sector demand is indicated by several salient features of the Zambian domestic economy, which make it an attractive destination for foreign direct investment. With a mixed economy and liberal policies towards private and foreign direct investment, Zambia continues to be a hub of foreign investments. Foreign direct investments are destined mainly for the mining sector but also enter the non-mining sector since the economy is liberalised. Zambia is also endowed with natural resources of numerous kinds that form a bedrock for various industrial types. These industries provide and maintain a viable domestic market via inter-firm trade linkages. According to Zambia Vision 2030 (2006: 10), the main industries include construction, transport, mining, manufacturing and agriculture, all of which flourish partly due to natural resource endowments of the country but are also likely to use cement as an input for their businesses.

Specific examination of cement production reveals that it is in the non-mining sector of Zambia. Available literature suggests five factors that are largely responsible for increased foreign investments especially in the non-mining sector of Zambia. According to Phiri (2011), these are urbanization, political stability, growth of real GDP, exchange rate depreciation and infrastructural development.

Growth in real GDP is a measure of the size of the domestic market, a key consideration for foreign-owned firms that target domestic sales as opposed to exportation. In the case of cement, its bulky nature, which makes transportation difficult, makes it logical that production is located in close proximity to the final users. It is also relatively 'simple' to produce, not requiring inputs that are too sophisticated (although bulky) or rare, and has a straightforward production process. If it wants to access a specific market, it thus makes sense for a foreign firm to be present in that market rather than exporting to it[41]. Moreover, besides enjoying a mean rate of 5% of GDP growth since 2005, Zambia's urbanization development has also been positive, a factor that has been found to attract foreign direct investments mostly of the non-

[41] This probably explains why as the Zambian cement market promises to grow, new FDIs have opted to locate in the country as opposed to probably exporting the product to Zambia

mining type. In Zambia, Phiri (2011: 42) finds that a 1% increase in the degree of urbanization results in 8.5 % rise in foreign direct investment inflows in the non-mining sector.

A notable attendant correlate of urbanization in developing economies is housing shortage, which is also typical of Zambia. Zambia is faced with a critical shortage of housing and an enormous housing backlog currently estimated at about one million units (Zambia Government, 2006: 23). Urbanization growth coupled with housing shortages ensures domestic market availability for cement-producing firms like Lafarge to flourish. However, the role played by growth in domestic demand in enabling firms like LCZ to perform exceptionally well may not be peculiar to the Zambian economy. It might be reflective of the production and consumption dynamics that generally characterise the global cement industry.

> *"...globally, many of the producing nations utilize their cement for internal consumption within the growing local market and export markets tend to be regional in cement trade, but with significant variance in country concentration relative to local production"* (Selim & Salem, 2010: 3)

Such global dynamics probably also influence the locational decisions of cement producing firms like Lafarge. With forecasted growth in regional demand as noted earlier, expansion in SSA is probably what made economic sense and informed the locational decision of Lafarge.

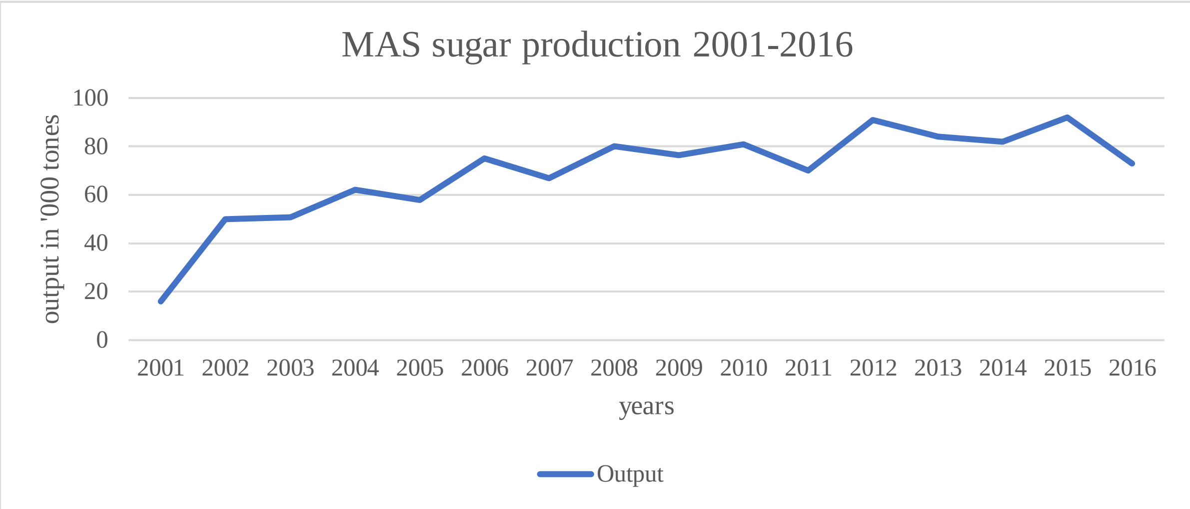
3.4.3 *Maragra Acúcar SARL (MAS)*

MAS is a sugar-producing firm in Mozambique. It is 76% foreign-owned by Illovo Sugar Group, which is headquartered in South Africa. MAS's ownership status provides another study case that is at least comparable to firms in Cluster 3 of the previous sub-section. This firm, however, provides a case of a foreign-owned investment that is different in terms of origin. Foreign ownership acquisition in this case is by a foreign investing firm from a similarly developing economy (South-to-South foreign direct investment) as opposed to the previous cases where investors originated from the developed economies i.e. North to South foreign direct investments. The choice gives chance to explore heterogeneities if any. Besides, MAS is an enterprise dealing in a product, which is relatively technologically less intensive, and relative to cement, can easily be exported.

MAS was originally known as Maragra Marracuene Açucar; and was founded by the Portuguese Petiz family in 1968. Sugar production, however, started in 1970. Initially the firm performed well; and at its peak in 1972, produced 44 100 tons of raw sugar (Sutton, 2014: 82).

At the eve of the country's independence, however, the firm was nationalised. Just as in other SSA economies where nationalisation policies were embraced, such as Ghana, Uganda and Zambia, flaws in public sector enterprises' management started to emerge. In the case of Maragra Marracuene Açucar, these flaws led to total termination of sugar production operations at the firm in 1984. In the late 1980s, Mozambique adopted liberal reform policies, probably to tackle the nightmare of public sector enterprise management and resuscitate the economy. In 1992, Maragra Marracuene Açucar was denationalised and re-acquired by the original owners. In order to revive the firm's production operations, the original owners sold 50% of the stake to Illovo Sugar Limited (ISL). This was later followed by a rebranding and consequent birth of the current Maragra Açúcar SARL in 1997.

Just as in the previous case studies, foreign acquisition was associated with significant benefits to the acquired enterprise. In the case of MAS, an expansion programme that catapulted the enterprise's productive capacity and general performance followed acquisition of part of the stake in the firm by Illovo. The programme undertaken in 1999 entailed expansion of cane plantation acreage to a magnitude of 4000 hectares and building of a new factory worth US$55 million (Sutton, 2014: 82). However, a flooding disaster that destroyed the entire plantation a year later prompted ISL to invest an additional US$ 15 million. This enabled the re-establishment of the plantations in 2001 with extra acreage which totalled 6 300 hectares (Sutton, 2014: 83). This investment expansion attests to the resilience of foreign-owned investments to both financial and non-financial shocks alluded to earlier. This investment programme translated into increased output by the enterprise in terms of both cane and sugar. In 2001, the firm processed 139 000 tons of cane and outputted 16 000 tons of sugar. In the following year, the firm produced 412 000 tons of cane and 50 000 tons of sugar respectively in 2002 (Sutton, 2014: 84). In terms of sugar production, the firm registered an over 200% increase. Within only two years, the firm had surpassed the previous record (44 100 tons) of sugar produced by the original enterprise before acquisition by Illovo. In figure 3.10, we notice that after rebuilding the infrastructure, sugar production at the firm took on a positive trend, although with some volatility as expected.



Source: Author's illustration based on figures from ISL integrated annual reports

Figure 3. 10: Annual production of sugar by MAS, 2001–2016

The performance of MAS illustrated in Figure 3.10, and which improved almost exponentially after Illovo's acquisition, can be associated with several factors. Just as in the previous cases, such factors may range from natural factors, like climatic conditions that favour cane growing in Mozambique, economic policy, and market conditions, along with firm-specific features enhanced by Illovo Sugar group's strengths above all. As an agro-based enterprise, the availability of favourable climatic conditions for cane growing in Mozambique was favourable. A review of ISL reports over the past fifteen years reveals the role played by climatic conditions in cane yields and, consequently, tonnage of sugar produced. Deviation between tonnage of cane and sugar produced is accounted for by the sucrose levels in the cane. The sucrose levels in the cane are in turn contingent on the climatic and weather conditions that characterise a particular season.

In terms of market and economic policy conditions, Mozambique's openness crucially affected the exportation capacity of firms like MAS and consequently its productivity. Coupled with bilateral and multilateral international trade agreements with the EU, US, COMESA and SADC countries, Mozambique provides prospective foreign investments with a wide international market. Under such enabling conditions, export-oriented firms like MAS have a good chance of flourishing.

However, natural and market conditions discussed above are not limited to either foreign or locally owned enterprises. Assuming that any firm irrespective of ownership status can exploit or suffer natural gains or disasters and the economic conditions in the country, then firm-specific characteristics can be viewed as partly explaining productive premiums of one firm, like MAS or a group of firms (foreign-owned firms for instance) over others, especially in the same industry, or even across industries. This is truer when comparing firms that have foreign ownership vis-à-vis domestically-owned firms as previous results tend to suggest. In the case of MAS, ISL, the acquiring firm is globally, a leading, low-cost sugar producer. It is also a noteworthy producer of high-value downstream products. The group has extensive operations in South Africa and operates a beet sugar factory in the US (Illovo Sugar ltd, 2002: 2). Since it can leverage successes elsewhere including the US, the group's economic and technical potency easily overcomes any impediments to production, marketing and related aspects in its affiliates. To some extent MAS' performance improvement after foreign acquisition shows how foreign-owned firms generally have similar superiority features vis-à-vis domestically-owned firms, irrespective of the origin of the foreign investment.

3.4.4 *Kamal Steel Limited (KSL)*

KSL is a steel manufacturing firm in Tanzania. This enterprise is entirely owned by Gagan Gupta, an Indian investor. Its ownership nature differs from the earlier case studies, which are partially foreign-owned. This distinction makes KSL another suitable case study for examination as an enterprise comparable to firms in Cluster 4.

Gagan Gupta set up KSL in 2004, with full production commencing in 2005. Gupta was an Indian tourist who had come to spend a family vacation in Tanzania and decided to set up the firm after a meeting with a Mr Salgar, which took the form of a detailed enquiry about the steel industry in Tanzania. The resulting enterprise, which turned out to be a horizontal foreign-owned investment, was a typical reflection of innovative entrepreneurship mind and skills of Gupta:

> *"It was my first trip to Africa. Then as a businessperson, I became interested in steel because we are involved with steel in India. I went to some individuals who were dealing with steel here in Tanzania and asked what they were doing and what*

the prices were like. I spotted a large gap between the raw material and the finished product"[42]

Besides entrepreneurial instincts, Gupta had prior hands-on experience in steel production. He started as a steel stockholder in the Agra area in India under the name Kamal Enterprises in 1993. In 1997, he decided to join India's manufacturing sector, setting up a steel furnace to produce steel from recycled scrap (Sutton & Olomi, 2012). In the 1990s, Kamal Enterprises became the first firm in India to manufacture steel from sponge iron. By the time Gupta started his Tanzanian venture, Kamal Enterprises was a deep-rooted steel producer in India (Sutton & Olomi, 2012).

Probably in affirmation of foreign investments' features, KSL grew rapidly after it commenced production in 2005. In 2007, KSL launched its second steel mill and, a year later, an air separation plant was set up, which enabled the firm to produce its own oxygen for use in production processes. By 2012, the firm was already under the process of setting up a new greenfield operation, which was to catapult its production twenty-fold, from 40 000 metric tons yearly to 700 000 metric tons at a magnitude of US$200 million (Sutton & Olomi, 2012). In 2016, the firm diversified by launching a TZ Shillings 10 billion factory for refining furnace oil in Zinga-Kerege. This is litmus to the growth of the firm over the years since its inception. The firm has also diversified into production of industrial and medical gases and power generation since its establishment in Tanzania. KSL currently employs over 1 000 workers and, by 2012, it had an annual turnover of US$10 million.

Unlike in the previous cases, the rapid growth and performance of KSL can be attributed to firm-specific features and economic conditions in Tanzania only. It is worth noting Gupta's prior experience in the steel industry. As indicated earlier, Gupta was a well-established player in the steel industry of India before setting up KSL in Tanzania. From a management perspective, this experience must have given the firm a competitive advantage over local rivals and possibly other foreign-owned firms in the same industry.

---

[42] https://marcopolis.net /leading-steel-manufacturer-in-tanzania-talks-about-business-and-economy.htm

Two host-economy features can be cited as probable contributors to KSL's exponential growth and performance: (i) the growing domestic market and rapid infrastructural development that create demand for firms like KSL and (ii) proximity to the Indian Ocean, which makes access to the rest of the world easy for exportation.

Throughout the four case studies, it is noticeable that:

(i) At least two types of foreign investments typify sub-Saharan Africa;

(ii) Most partially foreign-owned firms have a similar history of ownership dynamics; and findings do not agree with the cherry-picking effect. This is, however, not to dispel my earlier conclusions. This is because of the fact that almost all selected cases were formerly state-owned enterprises which were marred by public mismanagement in most SSA economies shortly after independence. This is what makes acquisition to be more amenable to negative selection bias.

(iii) Acquisition of part of the stake by foreign investors in the originally domestically-owned enterprises improved the performance of the acquired enterprises.

Additionally, at least three categories of factors may be said to account for the improved performance of acquired firms. These are; favourable host country economic policies, natural factors, and firm-level strength of acquiring foreign-owned firms. Above all, all the four case study firms are quite comparable to firms in clusters two, three and four in the previous sub-section. The clusters in question are characterised by higher levels of foreign ownership stakes, with Clusters 2 and 3 having some firms that are fully foreign-owned. These clusters are further composed of large firms, which are also capital-intensive, just as are the selected case studies.

GREL and MAS (Illovo) are two firms that reflect natural resource-seeking foreign investments, while LCZ and KSL tend to mirror domestic, market-seeking foreign investments. With the exception of KSL, all the cases studied are partially foreign-owned firms that were formerly either privately owned by local entrepreneurs or state-owned. The latter ownership was fuelled by nationalisation policies in host countries. GREL, LCZ, and MAS exhibited stellar performance after foreign take-over: they are capital-intensive, large firms that are export-oriented (except LCZ) and highly productive. Their jump in performance and characteristics exhibited attest to these cases' comparability to firms in the previous sub-section. Furthermore, their characteristics empirically mirror the results and assertions of cluster analysis in this thesis. In the case of GREL, LCZ, and MAS, the original firms were limping (before foreign acquisition) due to domestic management failures. The improvement in performance by LCZ and MAS in the early years of foreign acquisition is an empirical demonstration of foreign investments' superiority over locally-owned enterprises, a key hypothesis we alluded to in our cluster results.

Besides the natural factors and domestic economic policy reforms that probably favoured the performance of the cases examined above, it is worth mentioning key firm-level features of the foreign firms that acquired the then-limping local private- or state-owned enterprise. Lafarge, Illovo, and SIPH were all superior in their respective fields at the time of their acquisitions. They exhibited all the superior firm-level features, of which results are seen to be the key performance dimension along which foreign-owned firms and domestically-owned firms tend to differ systematically, according to the cluster analysis results discussed here. The superior features of these acquiring foreign firms, such as modern technologies, financial strength, and skilled manpower, which might have been transferred to the local subsidiaries, provided one mechanism through which such firms may have achieved improvements in performance.

3.5 Conclusions

In this chapter, we have again confirmed in the preliminary stages of analysis that foreign-owned investment firms and domestically-owned firms tend to systematically differ on a number of performance indicators using a multi-country broader data set. This is similar to findings not only in Chapter 2 but also to empirical findings of other studies conducted mostly in the developed world. Not only have most of these other studies been based on discriminant functional analytical methods where prior assumptions are made, mainly on the distribution of the data used, but, most importantly, the majority are country specific, like the analysis in Chapter 2 of this thesis.

In the second analytical stage, standard hierarchical clustering which was applied to group firms revealed intra-foreign-owned firms' classification, hugely driven by foreign ownership and export intensity. Statistical analysis of the clusters formed by the algorithm confirmed that groups with foreign-owned firms are associated with higher mean values than both the cluster with domestically-owned firms and the average for the whole data set. This confirms the systematic differences between the two types of enterprises as revealed at the first analytical stage. The similarity of this finding to those in Chapter 2 allows us to assert that previous findings in Chapter 2 are not peculiar to Uganda. They are likely to be a systematic phenomenon across foreign and domestically owned enterprises in SSA, and above all, it is basically firm-level features that matter (not country, economic group etc) in separating foreign and domestically owned investments.

The grouping of two foreign-owned firms in Cluster 1, a predominantly domestic firm cluster, is confirmation of the ability of our chosen clustering methodology in identifying firms based on the extent of their homogeneity as measured along the selected dimensions. The majority of these dimensions are performance-based, while some are more characteristic in nature. The ability of the clustering method is further supported by the composition of the outlier Cluster 5, with both domestic and foreign-owned firms, which are in all likelihood similar on the variable dimensions where data is available for the four outlying firms that constitute such a cluster. This implies that the choice of the method used was credible, given the empirical question that the chapter endeavoured to answer.

The relative importance of exportation in firm segmentation as revealed by the CART analysis at lower levels (not indicated at the tree diagram), tends to support the assertion that most foreign-owned investments are export oriented. As other studies have revealed findings indicated that foreign-owned firms export more on average compared to domestically-owned firms. Although some target either domestic markets, which usually crowds out domestic investments, most foreign-owned firms go for regional markets near host economies or international markets for export interests. Indeed, available empirical evidence confirms that foreign investments from the Asian sub-continent, especially Chinese firms, are attracted by bilateral trade agreements that SSA economies have with the Western world, notably AGOA and EBA (Henley *et al.*, 2008: 13). Such trade relations are exploited by foreign-owned investments to access export markets at favourable conditions, since SSA economies are constrained in meeting the demand in the export market opportunities created by the agreements. No wonder, although there are overlaps in terms of aim, such as market, resource and efficiency seeking, manufacturing foreign investments in SSA are majorly market-seeking, with key determinants being market size and potential (Chen *et al.*, 2015: v). Nevertheless, market-seeking foreign investments are more likely to be integrated in the domestic economy and assist local suppliers than other forms of foreign direct investments (Farole & Winkler, 2013).

Overall, results in this chapter agree with both findings in the previous chapter and findings from other studies that have used different methods and data sets. Specifically, and key to this chapter, is the confirmation that these differences are prevalent amongst foreign-owned and domestically-owned firms across SSA, and not unique to Uganda alone. Foreign ownership and exportation play a significant role in segmentation. As noted in the previous chapter, this exploratory multivariate analysis of data does not provide predictive mechanisms or establish causal effects of foreign ownership on performance, but paves the way for the same. By classifying firms into groups that can easily be studied relative to each other, discriminant predictive methods can be employed to further study the groups. Such causal mechanisms are further explored in the next chapter where the causal effect of foreign ownership on firm-level performance is investigated using sophisticated econometric methods and panel data.

One of the limitations of this analysis is related to data, the most visible of all being missing values just as in the previous chapter. The credible imputations performed have lessened the effect of missing data. Additionally, the study utilises UNIDO’s AIS 2010 survey data. The fact that these data were collected many years ago means that study results are subject to changes in the economic environment in SSA over the years. This implies that these study results ought to be viewed with caution but also that further research ought to be conducted using more recent data if it is available, if these analyses are to inform better policy formulation and probable implementation.

Chapter 4

Foreign Acquisition and Firm Performance in Sub-Saharan Africa

"*Evidence from Ghana*"

4.1 Introduction

In the previous chapters 2 and 3, this thesis provided evidence on probable[43] and considerable performance and characteristic premiums in favour of foreign-owned investments. This evidence was revealed in terms of several variables including but not limited to; labour productivity, employment, wages, capital intensity, output, and management experience. Additionally, this evidence is provided from both a specific country perspective and a multi-country scenario in Chapters 2 and 3 respectively. These findings so far point to, among other aspects, likely systematic differences and hence observed average performance premiums between foreign and domestically-owned firms in Uganda and SSA. Nonetheless, in the available empirical literature, it is a prevalent finding regarding ownership structure and firm performance, that firms with foreign ownership tend to perform relatively better than those with purely domestic ownership in typical host economies. In the empirical literature, this has been partly attributed to our previous evidence i.e. the systematic differences between foreign-owned and domestically-owned firms (Matthias & Javorcik, 2009: 45). What is until now inadequately known, however, especially in SSA, is whether the observed performance premiums and hence systematic differences between the two types of firms, are causally attributable to firms' ownership status (and to which extent) or the link is merely correlative as scholars like Navaretti *et.al* (2004) have claimed. It is clear in the evidence provided in the previous chapters that although foreign ownership is very important in segmenting the two types of firms into numerous clusters, it is also correlated with other variables. Moreover, whereas the observed systematic differences and associated performance premiums might be less surprising for Greenfield investments when compared to domestically-owned firms, foreign ownership arising out of acquisition of formerly domestically-owned firms highlight an empirical question regarding the extent to which the resulting performance may be ascribed

[43] In the previous chapters, I employ exploratory methods, whose results are largely descriptive and not necessarily causal.

to foreign acquisition itself. Whether foreign investors acquire firms that are high-performing in the first place (as chapter 3 findings tend to suggest) or whether foreign acquisition in itself positively and significantly affects the performance of acquired firms (and hence the observed premiums) is a subject that has been inadequately attended to by the previous chapters' analyses. It is the key empirical question, as indicated in chapter 1, which this chapter attempts to address. And given that part of the motivation was the scarcity of empirical investigations focusing on developing economies especially SSA, this chapter attends to this key issue by examining the effect of foreign acquisition of formerly domestically-owned firms on firm-level performance, providing such evidence from Ghana, a developing economy in SSA.

The issue of foreign-ownership premium has been widely studied, especially over the past decade. Studies have provided evidence on the effect of foreign acquisition on several firm performance outcomes. Some of the available studies in the empirical literature are; Benfratello and Sembenelli (2006), Aydin, Sayim & Yalaman (2007), Girma and Gorg (2007a), Huttunen (2007), Chari, Chen & Dominguez (2009), Matthias and Javorcik (2009), Guadalupe, Kuzmina & Thomas (2012), Fabling and Sanderson (2014), Weche Geluebcke (2015), Bentivogli and Mirenda (2016), and Shrivastav and Kalsie (2018). Most of these empirical studies confirm the existence of a positive effect of foreign ownership on for instance; wages, productivity, and R&D. Nevertheless, several other studies provide differing findings, indicating that a mixed picture of foreign-ownership effect on firm performance exists. For instance, whereas most of the abovementioned empirical studies find a positive and significant effect, Benfratello and Sembenelli (2006), using firms in Italy and UK, find that foreign acquisition does not generally lead to improved performance of acquired firms. Navaretti, Cooper & Venables (2004) stress that much of the available empirical evidence supports existence of a non-causal link but of a statistical association between foreign ownership and outcomes like productivity. This association is reflective of the evidence this thesis provides so far in chapters 2 and 3. Navaretti *et al* (2004) further state that in much of the causal studies undertaken, performance differences between foreign and domestically-owned firms, along variables like productivity, are less than in earlier estimations, and are frequently insignificant. Among firms in New Zealand, Fabling and Sanderson (2014) find positive effects of foreign acquisition on wages and output but no effects on productivity and capital intensity. In Germany, Weche Geluebcke (2015) finds a negative impact of foreign acquisition on employment and no productivity impact of the same on acquired firms. It is therefore clear that empirical findings still show little consensus regarding foreign acquisition effects on firm performance.

Besides the inconclusive findings in empirical literature so far highlighted, the majority of the studies have been conducted in the developed world or in developing economies such as in Asia and Latin America. Such analyses have hardly ever been based on economies in Africa, probably due to data limitations, among other impediments. Yet the continent has witnessed – (and continues to witness) – a surge in inflows of foreign direct investments in the last decade[44]. Elsewhere, majority studies have mainly focused on the link between foreign-ownership and productivity, neglecting analysis of other equally important aspects of firm performance like output growth, capital and skill intensity (Matthias & Javorcik, 2009). These are aspects of greater importance in developing economies like Ghana's and sub-Saharan Africa (SSA) as a whole, where inadequate domestic capital and higher levels of graduate unemployment, along with other socioeconomic problems, are witnessed. Up to this point, it is clear that it is not only the inadequacy of analyses in chapters 2 and 3, which justifies further inquiry, but also other aspects namely; inconclusiveness of available findings, biased focus on the developed world and limited scope on performance outcomes analysed so far by earlier studies.

Using a twelve-year-long panel data set on manufacturing firms, this chapter provides further empirical evidence on the causal link between foreign acquisition and firm performance. It provides this evidence from a developing economy in SSA, being Ghana, focusing on not only the usual outcomes of employment, productivity and wages but also on other crucial outcomes, like output, and skill and capital intensity. The analysis also looks, equally, at how capital levels are affected, despite this not being a direct firm-performance outcome.

This study aims to contribute to the growing literature on foreign ownership and firm performance in a number of ways. Firstly, the available data set enables an examination of the magnitude and direction of causal effects of foreign ownership on performance. In this endeavour, I have employed both traditional regression methods and matching methods with difference-in-differences estimation techniques, to address possible effects of inherent firm-level heterogeneity and bias due to selectivity on unobservables. Secondly, unlike many previous studies, this study does not focus only on productivity and wages, but also on other outcomes of firm operations, such as employment, output growth, capital, and skill intensity. These other outcomes can, potentially, be influenced by foreign acquisition. We also examine the effect of acquisition on one of the scarcest resources at firm level in the developing world:

[44] For instance, Ghana alone attracted 3.3 billion USD of FDI in 2017 (World Investment Report, June 2018).

capital. Finally, the study also provides evidence on the effects of foreign acquisition in the long run. The effects of foreign acquisition on outcomes like employment and wages may be delayed by costs associated with hiring and firing (Huttunen, 2007: 1). Equally, effects on outcomes like productivity and gross output may be delayed by within-plant structural or organisational changes. We cater for such delays in our analysis by examining the effects of foreign acquisition in numerous periods after acquisition in our robustness checks on initial results.

Besides the availability of data, Ghana's suitability for this case study is cemented by many other reasons. Ghana is one of the developing economies in SSA that shifted from a development strategy in which the government led the country's economic activities to one in which the private sector is at the helm. Due to the World Bank's conditionalities of the late 1980s and the failure of public sector economic management by post-colonial administration, Ghana put in place a divesture programme that saw the sale of government stake in state-owned enterprises (SoEs). Although launched before 1990, the divesture process was not successfully implemented until several measures were taken by government. These measures were implemented seriously in the first quarter of 1994 (Adda, 1996: 6), three years after the Regional Program for Enterprise Development (RPED) survey had started. Moreover, the process started with the sale of the least attractive SoEs, with those that attracted the highest number of investors being sold in the late 1990s (Potter, 2015: 6). This lends credence to my determination that the first acquired firms came into existence around 1994, as will be seen later. Foreign investors were highly encouraged to bid on SoE assets; and many enterprises were bought by foreigners (Potter, 2015: 8). This confirms that these policy changes exogenously shifted ownership of some firms from government or local parties to foreign investors.

The rest of this chapter proceeds as follows: In Section 4.2, I discuss briefly the theoretical guidelines to the study and review the existing empirical literature on foreign acquisition and firm performance. Section 4.3 describes the empirical strategy and results from regression analysis, while Section 4.4 specifically details the matching methods and associated empirical results that act as robustness checks for regression results. Section 4.5 presents the conclusions.

4.2 Theoretical and Empirical Literature

4.2.1 Theoretical basis

Analysis in this chapter is guided by two mutually reinforcing theoretical formulations concerning foreign ownership and firm performance. It is sufficient to recall that foreign ownership in this chapter is limited to ownership arising out of acquisition of a formerly domestically owned firm.

The first theoretical perspective is linked to the general hypothesis of ownership advantages[45] for foreign-owned firms over domestically-owned firms (Dunning, 1980; Hymer, 1960; Makoni, 2015).These advantages give foreign-owned firms a competitive edge over domestically-owned firms in host economies. These proprietary assets usually take the form of new technologies or patents, managerial expertise, superior access to capital, and product differentiation, among others. Once these are transferred to the subsidiary (acquired) firms, a foreign ownership premium (FOP) is generated, manifested in augmented performance of the subsidiary[46]. This hypothesis implies that the most productive firms within a specific industry or sector, are those that are more likely to engage in foreign investment through either setting up new plants or acquiring domestically-owned firms (Bentivogli & Mirenda, 2016: 2)[47]. In this study, testing this theoretical perspective requires the identification of the effect of foreign acquisition on the post-acquisition performance of the firm in question.

The second theoretical formulation relates the superior performance of foreign-owned firms to *ex-ante* selection bias. This bias comes from differences in micro-level features of firms that are acquired by foreign investors; and bias can either be a result of negative or positive selection. In instances of positive selection, foreign investors usually target high-performing domestic firms for acquisition, making their post-acquisition performance highly attributable to the selection process. Reviewed literature has referred to this as the “cherry-picking effect” (Rodríguez & Tello, 2014: 8), an effect that clustering results in chapter 3 also seem to reflect. The negative selection perspective posits that high-performing foreign firms may decide to acquire underachieving domestic firms and then remove their existing management in order to

---

[45] This theory that also guides analyses in the previous chapters is well detailed in chapter 2’s theoretical background section

[46] Bentivogli and Mirenda (2016) refer to this as ex-post forward linkages

[47] This is comparable to the operating efficiency theory in management literature, where acquisition is said to arise when the acquirer confirms complementarities with the targeted firm in terms of production operations. In this case, acquisition is likely to occur for those firms that are performing well. Post-acquisition performance in this case is contingent on the extent to which assumed complementarities exist in reality.

fully exploit the firm's potential[48] (Bentivogli & Mirenda, 2016: 2). Acquisition of underperforming domestic firms can also be the result of information asymmetry about the firm's performance. In this study's context, this theoretical perspective is tested by an examination of the pre-acquisition differences in performance and structural features between the acquired firms and those firms that remained purely domestic.

In order to fully understand whether foreign acquisition has a causal effect (FOP) on firm performance, I separated performance differences associated with pre-acquisition factors from those related to firm acquisition by foreigners. In this endeavour, I controlled for, among other aspects, endogeneity effects likely to arise from the selection process, and unobserved firm-level heterogeneity that could potentially influence the outcome of interest. This is what makes the two theoretical underpinnings conjointly reinforcing in this study.

4.2.2 Empirical Literature

The body of empirical literature on the causal nexus between foreign acquisition and firm-level performance has been growing over time and is now very topical. Among recent studies on the topic is Wang and Wang (2015a), who employ propensity score matching (PSM) and difference-in-differences (DID) to investigate the effect of foreign acquisition on local firms in China, with a focus on various performance outcomes. Their findings reveal that, relative to non-acquired firms, acquired firms showed a post-acquisition increase in real wages, a decrease in capital labour ratio, a 10% rise in employment, and gains in productivity.

Wang and Wang's (2015a) wage results uphold those of Girma and Gorg (2007a), who found positive wage effects of acquisition in the UK for firms acquired by US foreign investors. Girma and Gorg also employ a difference-in-differences matching approach in their analysis, with their findings revealing substantial heterogeneity in the acquisition effect on wages, chiefly associated with nationality of acquirers and workers' skill groups. Wage results mentioned above further corroborate findings by Conyon, Girma, Thompson & Wright (2002) who utilise panel data on firms in the UK for the period 1988– 1994, finding that acquired firms pay on average 3.4% more than similar domestic firms. They employ a difference-in-differences method implemented in a regression framework (Conyon *et al.*, 2002: 7).

[48] This is comparable to "the theory of managerial discipline" in management science, which has it that acquisition is a result of natural selection, in which inefficient firms are taken over by new shareholders and undergo restructuring to enhance their efficiency.

Using panel data for 1991–1998, Almeida (2007) employs regression and generalised difference-in-differences estimation methods to investigate the acquisition effect on wages. Almeida's findings show significant wage increments after takeovers among Portuguese firms. Employing both regression and PSM with DID on longitudinal data from Finnish firms, Huttunen (2007) finds a significant and positive effect of acquisition on wages in the second and third years after takeover. Similarly, in their analysis of foreign acquisition on New Zealand firms, Fabling and Sanderson (2014) find significant acquisition effects on wages. Acquired firms experience an average of 6 % to 8% increase in wages when compared to non-acquired firms over a three-year horizon (Fabling & Sanderson, 2014: 13). This scholarly evidence confirms the acquisition effects on firm wages. However, in line with the inherent inconclusiveness on this subject, still other studies have found differing results. In attempt to study the effect of foreign acquisition on wages and total factor productivity in the years following takeover, Bandick (2011) utilises firm-level data on Swedish firms for the period 1993–2002. In this study, the probable endogeneity of the acquisition decision is handled using a combination of instrumental variable techniques and propensity score matching with difference-in-differences estimation methods. After addressing firm heterogeneity aspects, the study finds no effects of foreign acquisition on wage growths of acquired firms (ibid.).

The productivity results from Wang and Wang (2015a) echo the findings of Matthias and Javorcik (2009) concerning Indonesian firms. Using longitudinal data from Survei Manufaktur, Matthias and Javorcik (2009) applied matching methods with difference-in-differences techniques and found statistically significant increments in the productivity of acquired firms relative to similar non-acquired firms. Specifically, their findings indicate that acquired firms on average post between 10 % and 13.5% more on productivity when compared to non-acquired firms that were similar in the pre-acquisition period (Matthias & Javorcik, 2009: 47). Productivity improvements are prevalent in the acquisition year and subsequent years; and the findings are robust to different measures of productivity. Elsewhere, using panel data for 1986–2008 on Hungarian firms, Earle, Telegdy & Antal (2012) employ regression and propensity score matching with difference-in-differences estimation methods to investigate the wage effects of foreign ownership. Matching on pre-acquisition data and controlling for fixed effects for firms and detailed worker groups, these scholars not only find an increase of between 12% and 28% on average wages but also that these wage effects tend to rise with potential enhancements in productivity (Earle *et al.*, 2012: 16). Their findings specifically indicate a 26.1% productivity premium for firms acquired by foreign investors (Earle *et al.*, 2012: 39).

Besides the mixed results on acquisition and productivity, where similar effects have been found, there are variations in the degree of such effects. Some other studies have found comparatively lower effects of acquisition on firm productivity. Bandick (2011: 14) study of Swedish firms using matching methods and instrumental variables reveals only 1%–and 2% difference in productivity growth between acquired and non-acquired firms five years after takeover. Some studies also find no effect at all on productivity after a change in firm ownership. Applying a generalised method of moments (GMM) system estimation method on Italian firms, (Benfratello & Sembenelli, 2006) study controls for unobserved heterogeneity, measurement errors and input simultaneity but finds no productivity effects linked to foreign ownership. Similarly, Fabling and Sanderson (2014) utilised matching and difference-in-differences approaches on New Zealand firms in an attempt to estimate the effect of foreign acquisition on firm performance. Their findings indicate that foreign acquisition has no significant effects on firm productivity. Combining matching methods and the difference-in-differences estimation approach, (Salis, 2008) examines acquisition effects on productivity among Slovenian firms. Utilising data on firms in manufacturing between 1994 and 1999, study findings do not show convincing evidence of positive effects of foreign acquisition on total factor productivity for acquired firms over the period considered (Salis, 2008: 1031). Findings from Salis's analysis echo those of Orazem and Vodopivec (2004), who employ regression methods to examine firm-level production efficiency in Slovenia for 1994–2001. Their findings indicate that changes in firm ownership have no impact on growth in total factor productivity (Orazem & Vodopivec, 2004: 24). It is therefore clear that acquisition effects on productivity also remain inconclusive, despite it being one of the most investigated performance outcomes at firm level. This implies that any further inquiry, as intended by this study, is essential.

Acquisition effects on employment in Wang and Wang's (2015a) findings converge well with several other empirical studies. Earle *et al.* (2012) uses panel data for Hungary for the period 1986–2008 and finds positive acquisition effects on employment of about 7% for acquired firms relative to non-acquired firms. These findings substantiate those of Lipsey's (2008) study of Indonesian firms. Using a panel of firms in the Indonesian manufacturing sector for the period 1975 to 1999, Lipsey employs regression and difference-in-differences with matching methods to examine employment growth after acquisition. Results indicate an annual 5.6 % growth rate of employment for foreign-acquired firms when compared to the always-private domestic firms (Lipsey, 2008: 16).

Post-acquisition growth in employment may be dependent on the path taken by new firm owners. If new ownership aims to create new sales opportunities and networks, a rise in employment is inevitable; however, if the acquisition is guided by managerial purging and restructuring, job losses are likely to occur (Fabling & Sanderson, 2014: 4). Notwithstanding, several other empirical findings reveal differing effects of foreign acquisition on firm employment. Employing a difference-in-differences methodology combined with propensity score matching, (Chari *et al.*, 2009) finds employment declining in acquired firms relative to non-acquired firms. Huttunen (2007) examines the effect of acquisition on employment and wages for employee skill groups in Finnish firms. Employing regression and propensity score matching methods on a panel spanning 1988–2001, findings indicate a significantly negative effect on employment for highly educated workers. Maioli, Gong and Yundan (2007: 24) examine the effect of foreign acquisition of SoEs in China using matching methods and difference-in-differences techniques. After controlling for growth, they find a contemporaneous negative effect on employment in acquired firms relative to those that did not change ownership. However, post-acquisition effects are found to be positive.

Relatively few studies have examined foreign acquisition effects on capital investments, output, and capital intensity as firm-level outcomes. One study that examines acquisition effects on output is by Matthias and Javorcik (2009). Their findings reveal positive effects on output. On average, acquired firms had sales more than 50% higher than in relatively similar firms that were not acquired. Increments in output are chiefly attributable to various restructuring processes that may entail new technologies and production infrastructure, which are usually transferred to the newly acquired subsidiary firm from the parent plants abroad. This is supported by the theoretical formulations of the internalisation theory motivated by Dunning and other scholars. Further investigation on acquisition effects on capital investments by Matthias and Javorcik (2009: 51) finds positive and significant average treatment effects.

The previous review of empirical literature on effects of foreign acquisition on firm performance highlights numerous important issues, which issues point to gaps that this study attempts to close. Evident in the review is the lack of consensus on the effect of foreign acquisition on studied firm-level performance outcomes. Whereas many studies find positive effects, some other empirical findings present differing results, especially regarding outcomes like; employment, wages, and productivity. This lack of consensus erases any doubts that further inquiry is necessary. It is further evident in the review that firm-level outcomes, especially output, capital, skill and capital intensity have received less attention in empirical investigations of the causal link between foreign acquisition and firm performance. These are equally important aspect of firm-level outcomes equally likely to be affected by foreign acquisition. Analysis in this chapter attempts to focus on these too. In the context of geographical orientation, to the best of my knowledge, virtually all studies have focused on either East Asian or on more developed Western economies. Focus on economies in Africa is scanty or non-existent so far, a gap that this analysis tries to close. Finally, in terms of methods, it is also glaring that application of matching methods and difference-in-differences estimation techniques is on the upswing in a bid to establish causal effects in the economics field[49]. In the context of this study, the method is appropriate more so as a robustness check, because foreign acquisition decisions are largely driven by firm-level features of targeted firms, evidencing "cherry-picking", an aspect that has to be handled carefully if causality is to be identified with certainty. This also lends credence to the suitability of the two theoretical orientations chosen to guide this chapter's analysis. Together they cater for pre-acquisition selectivity while meeting the endeavour to unmask post-acquisition causal effects of foreign ownership.

4.3 Empirical strategy

In the previous literature review section, regression techniques and matching with difference-in-differences estimation are some of the most popular methods that have been applied in studying effects of foreign acquisition on firm performance. Although these methods have been used to analyse firms mainly in the developed world, this chapter utilises the same methods for analysis of firms in SSA. Regression is used as the main technique of analysis while matching is employed for the purpose of robustness checks. These methods are detailed later in this section. In order to examine the effect of foreign acquisition on firm-level performance outcomes, I have utilised a 12-year-long panel data set. This panel has firm-level information covering the period 1991–2002 on a sample of 312 firms anchored in Ghana's manufacturing sector. Collected in six rounds by the World Bank and the Oxford Centre for the Study of African Economies under the RPED project, the sample is stratified by sub-sector, size, and location of firm. Foreign acquisition was identified through firms' responses to the disclosure statement, "Is this firm under exactly the same ownership and same legal status as it was when we visited in … (year)? If not, then describe these changes." If the description of the changes includes sale of some stake (at least 10%) to foreign investors, then the firm was taken to have

[49] Matching and DID methods have been popularly used in numerous fields like the medical or clinical studies. See studies like Staffa & Hill (2008), Pirracchio *et al.* (2012) and Zurakowski (2018).

been acquired. However, all the acquired firms had over 40% ownership stake taken over by foreigners. Firms' responses to this question were used to construct the acquisition variable which was a dummy. This implies that acquired firms were assumed to stay in same ownership conditions from the time of acquisition to the end of the study period. This meant that no reversals were allowed and effects of such in results estimations were not catered for. The way acquisition was identified and associated assumptions was dictated by the kind of data collection instrument used and the questions therein. Therefore, all other firms except those that were; (i) fully or partially state-owned, (ii) privately owned by local Ghanaians or (iii) domestically-owned by both Ghanaians and the state, were dropped. This was because it is firms with such structure of ownership that were expected to be acquired by foreigners. After dropping some of the firms, 138 firms remained, with 8.7% of these firms identified as having been acquired during the study period. I deemed this number of firms and the extent of acquisition, being 12 firms, feasible for analysis using my proposed methods[50]. As earlier mentioned, the divesture programme in Ghana had a slow start, and was characterised by very low sell-outs until late 1990s, when new measures were put in place.

On Table 4A in the appendix, the sample composition of these firms is shown. In terms of sector composition, the garment sub-sector posts the highest contribution at 21.9%, followed closely by metal and furniture. The small-scale, resource-intensive sub-sector (ssrii) is the smallest contributor at only 0.7%. More than half of the firms are located in Accra, followed by Kumasi with 37.7% of the firms in the sample.

In Figure 4.1, the treatment variation plots for the sample[51] are shown. The plots have firm identification numbers on the vertical axis and time (in years) on the horizontal axis. The upper panel shows treatment variation among large firms[52], while the lower panel shows treatment variation among small firms. The red patch shows firms that were acquired by foreign investors, according to their response to the statement outlined earlier. The blue patch relates to those firms that did not change ownership during the study period. From the panels, it is noticeable that most firms changing ownership were acquired in 1998, followed by those that changed ownership in 1994, with only two or three firms changing ownership between 1996

[50] See studies by Earle et al (2012) using 1.78% treated firms, Abadie *et al.* (2012) with 2.63% treated firms, Fabling and Sanderson (2012) with 0.3% treated firms, and Bentivogli and Mirenda (2016) with 0.17% treated firms.
[51] Generated using R package PanelMatch.
[52] Firms with log employment>= 1.698970

and 1997. This is more clearly revealed by the overall plot in Figure 4A in the appendix. It is also noticeable from Figure 4.1 that treatment is prevalent for the most part in large firms with only one small firm being among the treated. Prevalence of acquisition being more evident among large firms is yet another pointer to probable selection bias alluded to earlier. From the plots, it is further evident that there is a bigger pool of control firms to pick from for the matching stage of the analysis.

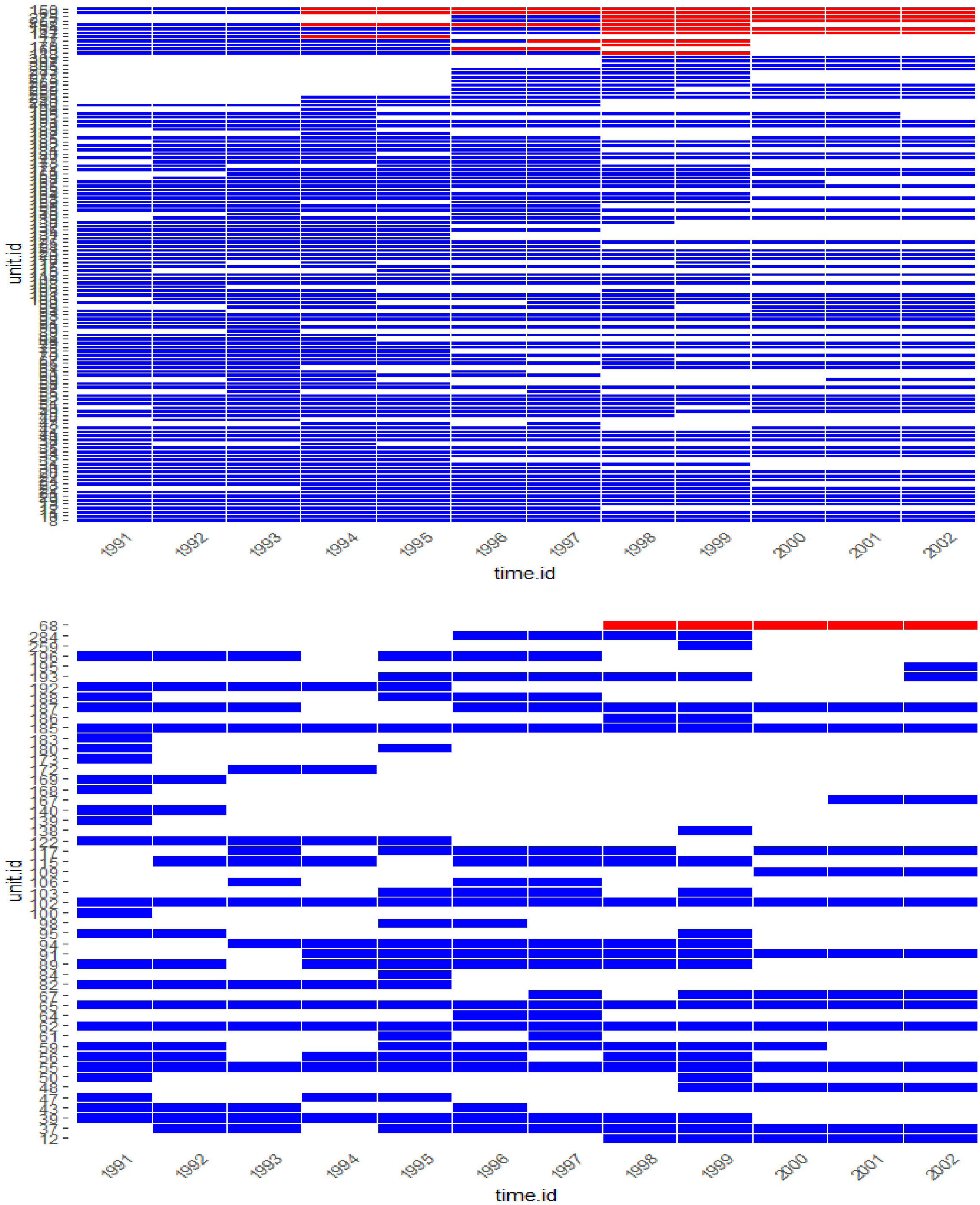


Figure 4. 1: Treatment variation plots for the derived sample

In Table 4.1, the descriptive panel summary statistics for the sample are shown comparatively for firms that were acquired by foreigners and those that remained domestic throughout[53]. Except for firm age and skill intensity, the rest of the variables are expressed in log form. In the last column, p-values are associated with the "raw" differences between the group means of the acquired and non-acquired firms in the sample[54]. It is evident that the "raw" means between acquired and domestic firms are significantly different for numerous variables.

Table 4. 1: Descriptive statistics for key variables for acquired and non-acquired firms

Variable	*Overall Mean*		*Overall Standard Deviation*		
	Acquired	Domestic	Acquired	Domestic	P>|t|
Employment	4.29	2.83	1.73	1.26	0.01
Wage	10.34	9.99	0.38	0.40	0.00
Productivity	14.74	13.89	1.60	1.25	0.13
Output	19.06	16. 47	3.13	1.94	0.03
Skill intensity	0.26	0.15	0.23	0.23	0.08
Capital	18.30	15. 92	3.05	2.74	0.00
Firm age	21.96	17.74	4.00	11.57	0.51

Source: Author's own computations based on sample data

In order to achieve my investigative objectives, I employed both regression and matching methods with difference-in-differences estimation techniques. Regression formed the main component of the empirical strategy while matching was used for robustness checks on the core results.

4.3.1 Regression methods

On Table 4.1, it can be noted that the descriptive statistics on key observable variables for the acquired firms and the non-acquired firms reveal differences between the two groups in the sample on most variables. This is an initial pointer to the probable selectivity regarding the foreign acquisition decisions of target firms. In most cases, foreign investors select their acquisition targets based not on chance but on specific criteria (Weche Geluebcke, 2015: 746). This lack of randomness in the foreign acquisition process of firms is of great concern when

[53] We generate these summary statistics using the "xtsum" stata command and report the overall means and standard deviations.

[54] Generated by "*reg outcomevar fdi, vce (cluster firm)"* to sidestep difficulties in conducting T- tests on group means in panel data.

estimating the causal effects of foreign acquisition on firm performance. I tried to control for any other observable and unobservable differences among firms by taking merit of the panel structure of the data. In the first stage of regression methodology, I estimated a pooled ordinary least squares (POLS) regression, which sets the standard for the separation of causal effects from the bias arising from selectivity and possible endogeneity. In the POLS, I regressed each key performance outcome against A_{it} , an acquisition dummy equal to 1 from the time a specific firm was acquired fully or partly by foreigners, and 0 otherwise. This dummy variable is 0 for firms that remain domestically-owned throughout the study period. I then included various controls and time dummies.

However, using the POLS does not solve problems associated with selection bias, possible heterogeneity and endogeneity issues. Among its limitations, the POLS disregard the space or individual firm effects. This model assumes that all firms are homogeneous in terms of dependent variables and that there are no other firm or industry effects on the regressands. Yet unobserved heterogeneity across firms, which may actually influence the outcome of interest can't be ruled out easily. As it may be recalled, although not explicitly revealed in chapter 2, cluster analysis of firms in SSA, Ghana inclusive, in chapter 3, revealed likely inherent heterogeneities amongst clusters with foreign-owned firms. The POLS further assumes that all slope coefficients of the X variables and intercepts are identical for all firms in the data set, which is unrealistic (Puron-cid, 2014: 205). Such unrealistic assumptions may bias estimates (for instance, of the variance for each of the estimated coefficients) leading to incorrect statistical tests and confidence intervals (Baltagi, 2005). The likelihood is thus high that the POLS may distort the true picture of the relationship between the regressor and the regressands across firms if the assumptions are not fulfilled.

I therefore conducted a second stage of regression analysis, in which I made the first attempt to sidestep selection, unobserved heterogeneity across firms, endogeneity issues, and the other limitations of POLS mentioned earlier. Endogeneity sources are myriad but mainly arise from all observable and unobservable time-invariant variables, which influence both the acquisition decision and the outcome variables of interest in this instance. Productivity differentials between foreign and domestic firms for example, partly account for higher wages in foreign-owned firms (Davies & Lyons, 1991). Omitted variables are also a realistic source of endogeneity. For instance, in this case, R&D orientation by foreign-owned firms influences both output and productivity, yet we do not cater for it in this analysis as no information about it - not even a qualifying proxy variable - is available. Besides, there are also possibilities of measurement errors, especially on wages and outright concealment on capital levels to avoid tax. Elsewhere, bias caused by simultaneity is equally possible. I note that, whereas foreign acquisition can, for instance, lead to higher levels of productivity among firms, highly productive firms might have a higher likelihood of being acquired by foreigners, hence positive selection. Yet negative selection is equally possible. In the second stage of the regression method, a fixed effects (FE) regression model was estimated for each performance outcome. One of the conventional ways of reducing the effects of selectivity is causal estimations, is by estimating a fixed effects regression model (Mummolo & Peterson, 2017). FE regression estimation also reduces problems associated with unobserved heterogeneity across firms in estimating causal effects. In a given panel data set, if units of analysis (like firms in this case) are likely to systematically differ from each other in ways that may not be observable, which differences may affect the outcome of interest, unit effects are often used to eliminate all between-unit variation so as to produce estimates of a variable's average effect within units over time (Wooldridge, 2010: 304). I therefore estimated a panel regression with firm fixed effects, which relates firm performance outcomes to acquisition status and controls. With slight modification, I adopted the following estimation equation as specified in Earle *et al.* (2012: 6):

$$y_{it} = \Psi A_{it} + \xi_t + \gamma_i + u_{it} \tag{4.0}$$

Where y_{it} is the natural log of a specific performance variable for firm i in year t, A_{it} is the acquisition dummy defined earlier. Ψ is the parameter of interest that captures the foreign acquisition effect on the outcome variable, while ξ_t denotes the 12-year effects. γ_i captures the firm fixed effects while u_{it} is the error term. I estimated equation (4.0) for each of the outcome variables of interest while excluding any other time-varying covariates that could potentially influence both foreign acquisition and the outcome variable. Additionally, I reported robust standard errors.

4.3.2 Empirical results from Regressions

In Table 4.2, the regression results for both the POLS and FE estimations are shown. The upper section of the table holds the results for the POLS estimations. For all estimates firm-level results show positive and significant coefficients, after controlling for sub-sectors and year effects. These results are subject to the limitations of the pooled regression mentioned earlier.

Table 4. 2: Effects of foreign acquisition on firm performance-estimates with OLS and FE

Full-sample ols	1	2	3	4	5	6
Variables	emp't	wage	productivity	output	capital	skill-int
fdi	1.37***	0.23***	0.70***	2.06***	2.63***	0.09***
	(0.174)	(0.045)	(0.159)	(0.301)	(0.265)	(0.025)
Constant	4.54***	10.15***	15.13***	19.14***	19.30***	0.51***
	(0.207)	(0.073)	(0.206)	(0.260)	(0.277)	(0.066)
Observations	1323	1500	1295	1297	1343	1210
Sector-dum	Y	Y	Y	Y	Y	Y
Year-dum	Y	Y	Y	Y	Y	Y
R^2	0.273	0.300	0.354	0.349	0.334	0.159
Full sample - FE						
fdi	0.094	0.099*	0.358*	0.433*	0.175	0.027
	(0.082)	(0.057)	(0.208)	(0.220)	(0.130)	(0.065)
Constant	2.84***	9.96***	13.75***	16.63***	15.77***	0.184***
	(0.049)	(0.028)	(0.085)	(0.084)	(0.038)	(0.020)
Observations	1338	1536	1310	1312	1358	1225
№ - firms (N)	134	138	131	138	133	137
Acquired firms	12	12	12	12	12	12
Sector dum	Y	Y	Y	Y	Y	Y
Year dum	Y	Y	Y	Y	Y	Y
R^2 - within	0.045	0.262	0.026	0.043	0.049	0.035

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Own results based on sample data

When a fixed effects model is estimated, the estimates are smaller. Results are, however, significant for wages, productivity, and output as observed in the lower part of the table. Specifically, it is evident that acquired firms tend to pay on average 9.9% more in wages when compared with non-acquired firms. The effects on wages corroborates the findings of Potter (2015: 8) in which workers in Ghana at former SoEs that were later privatised earned considerably more than their counterparts at similar SoEs that were not privatised.

Foreign-acquisition effects on wages may be explained by various theoretical channels including but not limited to efficiency wages to reduce shirking and labour turnover. Additionally, after acquisition, within-firm disputes associated with resistance to new changes may occur and new managers might respond to this my increasing average wages. Wage increments can as well arise from shared benefits from within-firm restructuring. Within-firm restructuring is even more realistic in this case given that most of the acquired firms were hitherto state-owned, in which restructuring is almost obvious after foreign takeover[55]. Irrespective of the channels, foreign acquisition effects on wages are usually associated with improvements in productivity in line with firm production theory. In this analysis, this is indicated by the estimated effects on productivity.

With reference to the results, acquired firms tend to be 35.8% more productive when compared to non-acquired firms. Acquisition effects on productivity are more than thrice the effects on wages, which is in line with the usual empirical interpretation that foreign acquisition stimulates productivity gains shared between employees and firm owners. Elsewhere, acquisition effects on output are even larger when compared to the rest of the outcome variables. It is noticeable that when compared to non-acquired firms, acquired firms tend to produce 43% more in output. The result on output corroborates earlier studies that were undertaken outside SSA such as those of Matthias and Javorcik (2009), Wang (2014), and Wang and Wang (2015b). Larger productivity and output effects can probably be attributed to the possibility of acquired Ghanaian firms, especially SoEs, being on a much lower boundary both technologically and organizationally prior to acquisition. This makes it relatively easier for foreign investors to raise productivity, output, and in the process also wages. This would be an interesting hypothesis to test but due to data limitations, this is not done in this thesis.

The positive (although insignificant) acquisition effects on capital levels are also worth noting. Positive acquisition effects are evident in terms of capital investments at 17.5% in favour of acquired firms. Acquisition effects on capital are indicative of the superiority of foreign-owned firms in terms of investment levels. This can be attributed to, among others, the usually less stringent conditionalities foreign-owned firms face in terms of access to finance and other forms of capital. The likelihood of foreign-owned firms accessing finance easily is equally acknowledged in Aliber’s theory discussed in chapter 2 of this thesis. Positive acquisition effects are also evident in terms skill intensity at 2.7% in favour of acquired firms. Similar results are noticeable in terms of employment, where acquired firms tend to have 9% more in employment when compared to their non-acquired counterparts. A probable explanation for the insignificant effects on employment is foreigners’ inclination to acquire already-large domestically-owned firms. On average, large firms generally grow more slowly than smaller

[55] Although within-firm restructuring is also possible in privately owned firms, it is more likely to occur in hitherto state-owned firms once these change ownerships.

firms (Karlsson, Lundin, Sjöholm & He, 2009: 188; Lipsey, 2008: 11); and hence it may not be surprising that growth in employment due to foreign takeover was not significant. In Ghana, during the divestiture programme, nearly all of the largest SoEs put up for sale were acquired by foreign investors (Potter, 2015: 6).

In Table 4D in the appendices, we control for the effects of previous firm performance by including lagged dependent variables in the estimations[56]. Results as seen from the table generally support evidence of acquisition effects on firm level performance outcomes especially on wages and output. Unlike in the main static model, this estimation also yields significant effects for employment and skill intensity. These results might, however, be problematic since the included lagged dependent variable is usually correlated with the error term, even if for instance OLS fixed effects estimators are used. This usually leads analysts to either turn to instrumental variables (that are difficult to find) or use of estimation methods like system GMM to handle the resulting endogeneity concerns. In this analysis I opted to instead perform robustness checks on the results using matching and DID which handles endogeneity.

The dissimilarity between the two sets (pooled OLS & FE) of results is usually an indicator of the degree of selection bias (cherry-picking) in the acquisition of firms by foreign investors whose decisions may also be based on the firm's inherent time-invariant heterogeneities (Earle *et al.*, 2012). In this thesis' context, it is clear that the FE estimation greatly tackles selection issues and provides more reliable results than the pooled OLS. Apart from employment, capital, and skill intensity, statistically significant acquisition effects on the rest of the variables are substantial. These results confirm that foreign acquisition tends to account for improvements in firm-level performance. These results therefore, answer the second key empirical question of this thesis by confirming that to a significant extent, firm level performance can be ascribed to ownership status. In this case specifically, results indicate that foreign-owned firms tend to perform relatively better than domestically-owned firms and this is partly due to differences in ownership. The size of the acquisition effects observed in Table 4.2 is relatively similar or slightly lower from empirical findings in the literature. Javorcik *et al.* (2009) found a 49 percent, foreign-output differential. For productivity, Waldkirch (2014: 26) found a 51 percent differential. In Earle *et al.* (2012: 10), findings indicated a 16 percent

[56] I estimated an equation of the form; $y_{it} = \lambda y_{it-1} + \Psi A_{it} + \xi_t + \gamma_i + u_{it}$ where λ is the effect of previous firm performance along the outcome variable of interest and the rest of the elements are as defined in equation (4.0)

wage differential in favour of foreign-owned firms using linked employer-employee data and an FE regression estimation method. In the next section I detail and justify the application of matching and DID estimation. Then I generate estimates of foreign acquisition to establish whether results that are at least to some extent reflective of regression estimates can be got. Through this estimation method, endogeneity is tackled and long run acquisition effects are also established.

4.4 Matching and difference-in-differences estimation

We earlier noticed from Table 4.1 that firms in the sample with foreign acquisition differ from those that are purely domestically-owned. Elsewhere, empirical evidence has shown that firms with foreign ownership usually pay, on average, higher wages and are also export-oriented (Matthias & Javorcik, 2009: 45). Because these differences do not give insights into whether the observed relatively superior performance of foreign-owned firms compared to domestically-owned firms is partly due to foreign ownership itself or otherwise, I estimated both POLS and FE regression models in an attempt to provide such evidence. Using foreign acquisition and firm performance as the independent and dependent variables respectively, results indicated that foreign ownership positively affects firm-level performance. The FE model results indicated that foreign acquisition positively influences wages, productivity, output, employment, capital, and skill intensity. These positive effects are significant for wages, productivity, and output. FE was employed because of the inherent limitations of the POLS, specifically regarding selectivity, and probable unobserved heterogeneity across firms, which may affect outcomes of interest. However, these regression models also have some limitations in estimating causal effects. FE for instance may not be a panacea for all selectivity problems (McManus, 2011: 19) and does not handle endogeneity problems per se, a key source of unreliable results in causal econometric literature. The use of FE trades consistency for efficiency, since it uses within-firm change, ignoring between-firm variations[57]. Additionally, from a broader perspective, under the FE estimation framework, firms act as their own controls. This implies that firms that change ownership largely drive the observed coefficient on the acquisition variable[58]. Because of some of these limitations, it was deemed proper to apply other estimation techniques as a form of robustness check on the regression results. Given the

[57] This meant I could not assess the effect of variables that have little within-firm variation. Moreover, parameter estimates may be imprecise with large SEs.
[58] This made establishing a counterfactual scenario a little more enigmatic.

nature of the data and the empirical question under investigation, I employed a matching with difference-in-differences estimation method in a bid to establish whether acquisition effects at least reflective of those from the FE estimations could still be obtained. In causal econometric literature, matching enables the analyst to create a feasible control sample for estimation purposes by restricting the control sample to only those firms with relative similarity to acquired firms in terms of observable pre-acquisition features. Matching therefore, enabled me to establish a counterfactual scenario for the acquired firms in furtherance of the study's objective of establishing the average causal effects of foreign acquisition on firm performance. Matching coupled with difference-in-differences estimation would also solve the endogeneity issues explained earlier.

4.4.1 Modelling foreign acquisition of firms

One of the most crucial decisions in performing matching and successfully satisfying the assumptions that underpin this technique concerns the baseline covariates that are used in the matching model. Adequate identification and measurement of these covariates is essential (Tanner-Smith & Lipsey, 2014: 5). These covariates should satisfy two main properties according to these authors (page 6) i.e. (i) the covariate can independently predict any outcome variable of interest either directly or otherwise through a relationship with another covariate. This ensures that baseline differences between the conditions on that covariate would produce differences between the conditions on the outcome variable net of any effect of the acquisition and (ii) the covariate, or one with which it interacts, must differ between the conditions at baseline and measured with a good level of reliability. With no control over measurement, I handled identification in the inaugural stage of the matching strategy by empirically modelling the changeover process of potential target firms from local to foreign ownership. In this endeavour, I employed the probit model specified in (4.1), with a binary outcome variable of a firm being taken over by foreign investors.

$$P_r(A_{it} = 1|X_{it-1}) = \phi(X_{it-1}\beta + \varepsilon_{it-1}) \tag{4.1}$$

In (4.1), A_{it} is a dichotomous variable describing the acquisition status of firm *i* in year *t*. X_{it-1} is a vector of observable firm characteristics prior to acquisition and ϕ is a standard normal cumulative distribution function. In this model, I used the observable characteristics of prospective firms as predictor variables. The selection of the predictors included in (4.1) was guided by economic theory, the theoretical formulations explained earlier, and earlier studies that have employed matching with DID[59]. I lagged these predictors, except age, to ensure prediction is based on pre-acquisition performance of prospective firms. According to Matthias and Javorcik (2009: 45), potential foreign investors base their acquisition decisions hugely on basic observable information on targeted firms, for instance, the size of firms, capital levels, productivity, age, exportation status, among other characteristics.

In Table 4.3, the results from the probit model[60] are shown. Results indicate that firms that are relatively larger are more likely to be acquired by foreign investors. The coefficient on the lag employment covariate is positive and significant at 10%. Additionally, results indicate that firms that are relatively older, more productive and publicly owned are more likely to be acquired by foreign investors. The coefficients on these variables are positive and significant at 5% or 10%. Regarding public ownership, the significance is probably due to the policy shift that the Ghanaian government ushered in shortly after the initial waves of data collection. This saw some hitherto SoEs change ownership. The Ghanaian government also sold off its stake in some jointly owned enterprises to foreigners. Although a key predictor of foreign acquisition as indicated in numerous empirical studies, export status is nonetheless not significant in my data.

Table 4. 3: Probit results – Predicting foreign acquisitions

Variables	fdi	Std Error
lag employment	0.525*	0.319
lag wage	1.182**	0.565
lag productivity	0.448*	0.244
lag exports	0.430	0.400
lag public ownership	6.263***	1.601
Firm age	0.220***	0.025
Prob > chi^2	0.000	
Chi^2	223.54	
Pseudo R^2	0.440	

Note: *** p<0.01, ** p<0.05, * p<0.1

[59] See studies like; Weche Geluebcke (2015), Fabling and Sanderson (2013), Chari *et al.* (2009) and Huttunen (2007).

[60] I computed pseudo r^2 from (constant model LL - full model LL)/ constant model LL.

Given the results from the probit model, it was clear that use of such covariates at their baseline level would yield credible results from the matching process. As later details reveal, I used these covariates in the calculation of Mahalanobis distances, which were used in the refinement of the matched sets.

4.4.2 The Matching Strategy

In this analysis, I adopted a matching procedure for causal inference with Time Series Cross Sectional (TSCS) data as recently devised by Imai *et al.* (2019). In this chapter, I use the words 'treatment/treated' interchangeably with 'acquisition/acquired'. We recall that the data set consists of 138 (N) firms covered for 12 (T) years. For every firm indexed by $i = 1,2 \ldots N$, at year indexed by $t = 1, 2, \ldots T,$ I had an outcome variable of interest, y_{it}. This outcome variable in my case may be one of employment, mean wage, productivity or output. I also had a dichotomous acquisition (treatment) variable, A_{it}, which takes on the value 0 for firms that stay domestically owned throughout the 12 years and value 1 for firms that were acquired by foreigners and only from the period of acquisition to T. I also observe a vector of K time-varying covariates, z_{it}. In a specific year, I assumed that z_{it} is realised before A_{it}; and that outcome, y_{it} is realised after z_{it} and A_{it} have been realised. This direction of the assumed causal order ensures that z_{it} is not affected by the acquisition; and that the covariates therein take their values prior to the firm being acquired. Importantly, z_{it} can include lagged outcomes (Imbens & Wooldridge, 2007: 3).

4.4.2.1 Defining the Average Treatment Effect on the Acquired Firms

The average treatment effect on the treated (ATT) is one of the most commonly studied estimands in causal econometric literature (Imbens & Wooldridge, 2007: 4)[61]. In this instance, ATT is the average effect of foreign acquisition on the performance variables of those firms that were acquired during the study period[62]. Defining my ATT started with specifying non-negative integers F and L. The former represents the outcome of interest measured at F years after foreign acquisition of a firm. Given a firm acquired in 1994, F = 0 if our interest is in the average effect of acquisition on an outcome variable measured in 1994, an effect Imai *et al* (2019: 9) refers to as contemporaneous. F is set to 1 if one has an interest in measuring the ATT on an outcome of acquired firms one year after acquisition, whereas L specifies the number of lags to adjust for in the matching process.

---

[61] See other studies such as Heckman and Robb (1984) and Rubin (1977).
[62] A similar cardinal objective in Weche Geluebcke (2015) on acquisitions in Germany.

I initially defined my ATT of foreign acquisition based on the specifications below.

Let potential outcome under (not under) firm acquisition, P_a (P_0), be expressed as:

$$P_a = y_{i,t+F}\left(A_{it} = 1, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) \tag{4.2}$$

$$P_0 = y_{i,t+F}\left(A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) \tag{4.3}$$

Then

$$ATT(F, L) = E\{P_a - P_0 | A_{it} = 1, A_{i,t-1} = 0\} \tag{4.4}$$

In both (4.2) and (4.3), the acquisition history, $\{A_{i,t-l}\}_{l=2}^{L} = \{A_{i,t-2}, \cdots, A_{i,t-L}\}$, is set to the realised pre-acquisition period.

Referring to the above specifications, if a firm for instance was acquired ($A_{it} = 1$) in 1994 (t), and one sets F = 1, (4.2) would express the value of the potential outcome of interest in 1995 (t + F); while (4.3) would give the would-be (counterfactual) potential outcome had the firm remained domestically-owned i.e. $A_{it} = A_{i,t-1} = 0$. From (4.4), if L= 2, ATT (1,2) gives the ATT of foreign acquisition on the outcome of interest, for instance, average wage, one year after acquisition, while assuming that the possible outcome is dependent on the acquisition history for a probable two years back.

4.4.2.2 Key Identification Assumptions

In this analysis, identification of the effects is based on the following assumptions:

The choice of the values of F and L in this analysis implies the assumption that the potential outcome for firm i at time $t + F$ does not depend on the acquisition status of other firms, for example, $A_{i't'}$ with $i' \neq i$ and for any t' and does not also depend on the previous acquisition status of the same firm after L years, $\{A_{i,t-l}\}_{l=L+1}^{t-1}$. By implication, except for some carryover effects up to L years, there is no allowance for spillover effects.

Another key assumption in establishing causality using this matching method is that of unconfoundedness in treatment (acquisition) assignment. Specifically, in this analysis, given F, L and my estimand of interest, I assumed that conditional on the treatment, outcome, and covariate history up to $t - L$, the treatment assignment is unconfounded. This assumption is also referred to as sequential ignorability. In this study's context:

$$\left\{ y_{i,t+F}\left(A_{it} = 1, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right), y_{i,t+F}\left(A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) \right\} \perp\!\!\!\perp A_{it} | A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2,}^{L} \{y_{i,t-l}\}_{l=1,}^{L} \{Z_{i,t-l}\}_{l=0}^{L} \quad (4.5)$$

where Z_{it} is a vector of observed time-varying confounders for firm *i* at time period *t*. The assumption will be violated if unobserved confounders do exist. Additionally, if the treatment, outcome, and covariate histories before time $t - L$ confound the causal relationship between A_{it} and $y_{i,t+F}$, the assumption will be violated.

Together with another key assumption of overlap, assumption (4.5) forms the combined assumption of 'strong ignorability'[63]. This combined assumption implies that we can estimate the ATT by adjusting for differences in covariates between the acquired and control firms (Imbens, 2015).

Although the believability of the unconfoundedness assumption can be assessed[64], it is not directly testable (Imbens, 2015: 31; Imbens & Wooldridge, 2007: 2). In analytical applications with TSCS data, however, analysts are more concerned with unconfoundedness, which is likely to arise from unobserved variables (Imai *et al.*, 2019: 11). In this study therefore, I sidestepped the unconfoundedness assumption implied in (4.5) and employed a difference-in-differences (DID) approach. As proposed by Imai *et al* (2019), after conditioning on the acquisition, outcome and covariate histories, I adopted the parallel trends assumption, as specified in (4.6).

$$\mathrm{E}\left[y_{i,t+F}\left(A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) - y_{i,t-1} \,|A_{it} = 1, A_{i,t-1} = 0, \{A_{i,t-l}, y_{i,t-l}\}_{l=2,}^{L} \{Z_{i,t-l}\}_{l=0}^{L}\right] = \mathrm{E}\left[y_{i,t+F}\left(A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) - y_{i,t-1} \,|A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}, y_{i,t-l}\}_{l=2,}^{L} \{Z_{i,t-l}\}_{l=0}^{L}\right] \quad (4.6)$$

In (4.6), the conditioning set consists of lagged outcomes (except the immediate lag $y_{i,t-1}$), acquisition history and the covariate history.

4.4.2.3 Generating the Matched Sets

The next step in the matching process was to generate matched sets for acquired firms. For each acquired firm (*i*,*t*), I constructed a matched set M_{it}, of control firms that are identical to

[63] As conceptualized by Rosenbaum and Rubin (1983)

[64] For example, using lagged values of the outcome as pseudo outcomes and estimating the causal effect of the treatment on these pseudo outcomes.

that specific acquired firm in terms of pre-acquisition history from time $t - L$ to $t - 1$. The matched set is defined as:

$$M_{it} = \{i' : i' \neq i, A_{i't} = 0, A_{i't'} = A_{it'} \text{ for all } t' = t-1, \dots \dots, t-L\} \quad (4.7)$$

for the acquired firms with A_{it} = 1 and $A_{i,t-1}$ = 0.

It can happen that some acquired firms do not have control firms with which they share identical treatment history i.e. acquired firms for which $|M_{it}|$ = 0. For purposes of maintaining internal validity, such acquired/treated firms are excluded from the subsequent analysis. The matched set in (4.7) only adjusts for the acquisition history of the firms. But the parallel trends assumption in (4.6) requires further adjustments for other confounders, such as previous outcomes and probably time-varying covariates. This process of performing further adjustments is what Imai *et al.* (2019) refer to as the refinement of the matched sets. Refinement, once undertaken, partly sets the stage for future confirmation of the parallel trend assumption specified in (4.6). In this analysis, I used the Mahalanobis distance[65] in refining the matched sets. A description of this refinement is as follows;

After matching each acquired (treated) firm with at most ξ control firms from the matched set with replacement; thus,$|M_{it}| \leq \xi$, the average Mahalanobis distance (MD) between the treated (acquired) firm and each control firm over time is calculated as;

$$MD_{it}(i') = \frac{1}{L} \sum_{l=1}^{L} \sqrt{\left(v_{i,t-l} - v_{i',t-l}\right)^{\top} \Sigma_{i,t-l}^{-1} \left(v_{i,t-l} - v_{i',t-l}\right)} \quad (4.8)$$

for a matched control firm $i' \in M_{it}$, where $v_{it'} = \left(y_{it'}, z_{i,t'+1}^{\top}\right)^{\top}$ and $\Sigma_{it'}$ is the sample covariance matrix of $v_{it'}$. That is, given a control firm in the matched set, the standardized distance is computed using the lagged outcome variable and covariates and it is averaged across time periods.

Once the distance measure $MD_{it}(i')$ is computed for all control firms in the matched set, then the matched set is refined by selecting up to ξ most similar control firms that satisfy a specified caliper constraint C and we give zero weight to the other matched control firms. This allows us to choose a subset of firms from the original matched set, that are most similar to the

[65] This metric has similarly been utilised in studies that have, as in this analysis, a combination of propensity score matching and DID, for example by Matthias and Javorcik (2009) and Chari *et al.* (2009).

acquired firm in terms of observed confounders. For the acquired firm (*i*,*t*), the refined matched set is defined as;

$$M_{it}^{*} = \left\{ i' : i' \epsilon M_{it}, MD_{it}(i') < C, MD_{it}(i') \leq MD_{it}^{(\xi)} \right\} \tag{4.9}$$

where $MD_{it}^{(\xi)}$ denotes the ξ^{th} order statistic of $MD_{it}(i')$ among the control firms in the original matched set M_{it}. Once the matched sets have been refined, the estimator can be employed to get the estimated ATT for the outcome variable of interest.

4.4.2.4 The difference-in-differences (DID) Estimator

After obtaining the refined matched sets, the ATT of changing firm ownership from domestic to foreign is estimated. This involves, for every acquired firm (*i*,*t*), the estimation of the counterfactual outcome, $y_{i,t+F}(A_{it} = 0, A_{i,t-1} = 0, A_{i,t-2}, \ldots\ldots, A_{i,t-L})$ using the weighted average of the control firms in the refined matched set. Then the difference-in-differences estimate of the ATT for every acquired firm is generated and averaged across all acquired firms. This estimate is formally specified as:

$$\hat{\lambda}(F, L) = \frac{1}{\sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it}} \sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it} \left\{ \left(y_{i,t+F} - y_{i,t-1} \right) - \sum_{i' \in M_{it}} w_{it}^{i'} \left(y_{i',t+F} - y_{i',t-1} \right) \right\} \tag{4.10}$$

Where: $D_{it} = A_{it}(1 - A_{i,t-1}) \cdot 1\{|M_{it}| > 0\}$; while $w_{it}^{i'}$ represents the non-negative normalised weight such that $w_{it}^{i'} \geq 0$ and $\sum_{i' \in M_{it}} w_{it}^{i'} = 1$.

$D_{it} = 1$ only if firm (*i*,*t*) changes the acquisition status from the control condition at year $t - 1$ to the acquired status at year *t*, and has at least one matched control firm.

In using matching and DID estimation for robustness checks on regression findings, one other aim was to establish acquisition effects in the long run. However, the specification of ATT in (4.4) does not specify the future treatment sequence. Because of this, the matched control firms might include those firms that get acquired before the outcome of interest is measured at time t+F. Likewise, some acquired firms might return to control conditions before time t+F. For instance, a foreign investor having been a victim of probably information asymmetry might withdraw from owning a stake in a domestic firm after realizing that the deal is not worthy as it had been thought prior to acquisition. In my case, however, the nature of data can’t allow me establish such reversals. In this chapter, the assumption and hence keen interest were in ATT of a stable policy change where the counterfactual scenario was that; (i) for an acquired firm, acquisition does not occur before the outcome is measured and (ii) for the acquired firms, foreign ownership will be in place for at least F time periods. Given that one of the motivations for choosing matching and DID estimation as my robustness check procedure was because of my desire to re-estimate acquisition effects and in a long run i.e. where F>0, I estimated an ATT of a stable policy change relative to no policy change among the acquired firms based on a re-defined specification as shown in (4.11).

$$\mathbb{E}\left[y_{i,t+F}\left(\{A_{i,t+l}\}_{l=1}^{F} = \mathbf{1}_F,\ A_{it} = 1, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) - y_{i,t+F}\left(\{A_{i,t+l}\}_{l=1}^{F} = \mathbf{0}_F,\ A_{it} = 0, A_{i,t-1} = 0, \{A_{i,t-l}\}_{l=2}^{L}\right) \middle| \{A_{i,t+l}\}_{l=1}^{F} = \mathbf{1}_F,\ A_{it} = 1, A_{i,t-1} = 0 \right] \tag{4.11}$$

Where $\mathbf{1}_F$ and $\mathbf{0}_F$ are F dimensional vectors of, respectively ones and zeros. The acquired (matched control) firms are those that stay under foreign ownership (control) conditions throughout F time periods after acquisition has taken place while the matched control firms are assumed not to get acquired by foreigners at least for F time periods after acquisitions have occurred.

$$M_{it} = \{i' : i' \neq i, A_{i't} = A_{i't+1} = \cdots = A_{i't+F} = 0, A_{i't'} = A_{it'} \text{ for all } t' = t-1, \ldots, t-L\} \tag{4.12}$$

To estimate the ATT based on (4.10), we utilize the idea of marginal structural models in order to make covariate adjustments while sidestepping post-treatment bias (Imai *et al.* 2019). Specifically, we first constrain the matched set for each acquired firm (i, t) such that the matched control firms do not get acquired at least after time t+F. Then the propensity score is estimated by modelling the treatment assignment, for instance, by means of a logistic regression as in (4.13).

$$e_{it}\left(\{U_{i,t-l}\}_{l=1}^{L}\right) = P_r\left(A_{it} = 1 | U_{i,t-1}, \cdots, U_{i,t-L}\right) = \frac{1}{1+exp\left(-\sum_{l=1}^{L} \beta_l^T U_{i,t-l}\right)} \tag{4.13}$$

Unlike the above setting, the model must be fit to all firms including those that are not in the matched sets in order to model the entire treatment sequence. Using the result from marginal structural models, the weights are then calculated as,

$$\omega_{it}^{i'} = \prod_{f=0}^{F} \frac{e_{i,t+f}\left(\{U_{i,t+f-l}\}_{l=1}^{L}\right)}{1-e_{i,t+f}\left(\{U_{i,t+f-l}\}_{l=1}^{L}\right)} \tag{4.14}$$

For $i' \in M_{it}$ and $\omega_{it}^{i'} = 0$ if $i' \notin M_{it}$. In the final step the DID estimator in (4.10) is employed to get an estimate of the long-term ATT under the specified treatment sequence as defined in equation (4.11).

4.4.2.5 Covariate Balance Checks

In a typical evaluation of any treatment, policy or intervention, the key question is not whether the treated and the untreated units should be compared, but rather which units should constitute the comparison such that they represent the treated had they not been treated (Imbens & Wooldridge, 2007: 7). In this study's case, the question was which non-acquired firms best characterised the acquired firms (essentially in the pre-acquisition period) such that a valid counterfactual scenario could be achieved to facilitate the desired comparison. In propensity score matching, a key test of the suitability of the intended comparison between treated and control observations is a balance check on the covariates used in the matching. In this framework, I examined the resulting covariate balance between the acquired and the matched control firms to get insights into whether the two sets of firms were indeed comparable with respect to observable confounding variables. I examined the Standardised Mean Difference (SMD) of each covariate between the acquired and control firms in the pre-acquisition period. The mean difference was standardised at a given pre-acquisition period using the standard deviation of each covariate across all acquired firms in the data.

Given an acquired firm (i,t) with $D_{it} = 1$, the covariate balance for variable k at the pre-acquisition period $t - l$ was defined as:

$$B_{it}(k,l) = \frac{v_{i,t-l,k} - \Sigma_{i' \in M_{it}} w_{it}^{i'} v_{i',t-l,k}}{\sqrt{\frac{1}{N_1 - 1} \sum_{i'=1}^{N} \sum_{t'=L+1}^{T-F} D_{it'} \left[v_{i',t'-l,k} - \bar{v}_{t'-l,k} \right]^2}} \tag{4.14}$$

Where: $N_1 = \sum_{i'=1}^{N} \sum_{t'=L+1}^{T-F} D_{i't'}$ is the total number of the acquired firms. This covariate balance measure was then further aggregated across all the acquired firms for every covariate and pre-acquisition time period as follows:

$$\bar{B}_{it}(k,l) = \frac{1}{N_1} \sum_{i=1}^{N} \sum_{t=L+1}^{T-F} D_{it} B_{it}(k,l) \tag{4.15}$$

4.4.2.6 Results from the Matching

Matching on two (L = 2) lags and using Mahalanobis distance to refine the matched sets, I estimated the ATT of foreign acquisition on firm performance for three (F = 3) years. The choice of F was crucial in this analysis. There are numerous explanations as to why foreign acquisition may not affect firm performance outcomes as soon as acquisition takes place[66]. It is, for example, realistic that changes in firms' mean wages are usually linked to changes in the composition of the workforce (Huttunen, 2007: 500). Such changes in workforce composition come with costs of adjustment (Hamermesh, 1988). This implies that the changes may not be immediate and hence wages may as well not change immediately. Therefore, even after foreign acquisition, effects of the same on wages and employment may take some time to manifest[67]. However, employment adjustment in foreign-owned firms is usually faster than in purely domestically-owned firms (Barba Navaretti, Checchi & Turrini, 2002: 4).

Additionally, any firm acquisition, be it foreign or domestic, is associated with internal post-acquisition reorganisation[68]. This may entail introduction of new work practices, on-the-job training, new technologies and new production processes. This further implies that changes in variables such as mean wages, output, employment, and other performance outcomes may not occur immediately after acquisition[69]. Based on the above, I decided to set my F equal to at most three years after foreign acquisition of firms. The effect may well be contemporaneous. For instance, immediately after acquisition, new management may increase wages or stop some production processes, leading to reduced output and other attendant effects.

On Table 4B in the appendix, the potential matched sets before refinement are shown. The first column contains the identification numbers for firms that were identified as having switched ownership, while the second column shows the respective years of this switch. In the third column, the number of control firms in the matched sets is shown. The variation in the number of control firms between 1998 switchers and 1994 switchers is because 1994 switchers can also use 1998 switchers as controls in the years before 1998. Overall, it is evident that the size of the matched sets was good enough to permit comparison.

In Table 4.4, are the weighted DID estimates of the ATT effect of foreign acquisition, using Mahalanobis distance in the matching. Results show that, on average and a year after acquisition, acquired firms realised increments in employment, wages, and capital compared to relatively similar domestically-owned firms. When one considers all the outcome years, it is

[66] A reason why considering $F \geq 0$ was deemed suitable in my analysis.
[67] Huttunen (2007:507) actually confirms that the effect might not be immediate.
[68] See Gbettor *et al.* (2013:2442) in their study on a Ghanaian telecom firm acquired by Vodafone.
[69] Moreover, due to measurement errors, the exact time of acquisition might be uncertain.

evident that acquired firms on average experienced a gradual rise in employment, wages, productivity, capital and output relative to similar domestic firms. This overall positive result is clearly visible in Figure 4.2, in which the ATT is plotted over time for some outcome variables of interest. These average treatment effects are statistically significant in terms of average wages and capital outlay in the second year after acquisition. Also noticeable is that employment effects are averagely positive but not significant just like in the FE estimation results. These results are generally reflective of the FE estimation results discussed earlier with slight differences visible for effects on output and productivity. These slight differences could be due to the fact that the two methods are somehow different and FE estimation is stronger and more precise than the matching method. Matching has been said to be weak since it is based on observable characteristics only. In the following discussion, however, I give a brief specific discussion of results from matching.

Matching results show that acquired firms realised statistically significant increases in wages by an average of over 25 percentage points when compared to parallel non-acquired firms, especially in the first two years after acquisition. In the year of acquisition, the increase (contemporaneous effect) was 31.8% this being statistically significant. Similarly, when compared to similar non-acquired firms, acquired firms posted over 19% increments in capital levels in year two after acquisition, a result that is also statistically significant. However, in year one, the effect is not significant and exactly mirrors the effect obtained using FE estimation earlier on. Wage effects although much larger, are reflective of those results obtained by the FE estimation. In the context of specifically wage increments at firm level, and given that these were average wages measured as a total wage bill relative to employment, numerous mechanisms can be cited through which such (larger than FE results) wage increments could occur. Two such mechanisms would be; the same people remaining employed but experiencing an increase in wages; and a firm altering the mix of workers leading to fewer low-paid workers and higher-than-average paid workers.

It is also shown that foreign acquisition has a positive ATT effect on productivity. On average, acquired firms posted over 10% increases in productivity relative to similar firms that remained domestically-owned over the study period. Although quite lower and insignificant unlike in the FE estimation, empirical studies employing similar methods of analysis, like Wang and Wang (2015b), find similar results regarding foreign acquisition and productivity. Nevertheless, statistically significant ATT on wages that are not accompanied by relatively similar effects on productivity pose an empirical contradiction regarding conventional firm production literature.

But this might occur due to numerous reasons. For instance, in Table 4.4, DID estimates indicate that employment increases, and that productivity also takes a positive trend in later years. This is may be suggestive of the fact that firms may be adding highly skilled or productive workers who may be equally well remunerated, but that it takes some time for all this to translate into increased output and productivity.

In the context of purely firm ownership, Fabling and Sanderson (2014: 12) attempt to motivate why foreign-owned firms may pay higher wages even when productivity remains unchanged. Explanations given include but may not be limited to foreign owners potentially securing workers from outside the country in question (Ghana), who may actually demand higher remuneration than domestic workers. Additionally, new firm owners may want to avoid the spillover of their superior technologies and vital information to rival firms. One of the conduits for such spillovers is through workers who leave the firm in question for a rival. Therefore, to prevent such turnover, workers may be given higher wages for the same level or for a less-than-proportionate increase in productivity. As hinted earlier in the discussion of FE results, foreign acquisition may result in the internal reorganisation of the acquired firm in terms of production and work practices, which might draw worker resistance and breed disputes. Managers might use increased wages to lessen such firm disputes regardless of the level of productivity (Conyon *et al.*, 2002). Foreign-owned firms have also been known to offer more productive on-the-job training to workers. And if this is the case after acquisition, then workers in acquired firms are likely to have a steeper wage profile and to thus acquire a premium with time (Görg, Strobl & Walsh, 2007). This premium is empirically echoed by Konings and Vanormelingen (2015), who examine the impact of training on wages in Belgian firms. Using an unbalanced panel for the period 1997–2006, Konings and Vanormelingen (2015) estimate firm-level wage equations based on Mincer's (1974) framework, finding a 12% wage premium for trained workers relative to untrained employees. The measurement of labour productivity may also be a solid explanation of the puzzle[70]. Elsewhere, rent-sharing across borders could account for the puzzle above. Wages in the acquired firm may to a great extent be linked to profits in the parent company abroad (Girma & Görg, 2007b: 100). This means that, at the same

[70] Foreign owners may practise transfer pricing by, say, reporting lower values of manufactured output going to parent firms outside Ghana, thereby reducing the measured value added. Under such circumstances, measurement results for productivity will be low, *ceteris paribus*.

productivity level, wages in the newly-acquired plant may significantly increase after acquisition.

Table 4. 4: Weighted difference-in-differences estimates of ATT with Mahalanobis distance

Variables	***F = 0***	***F = 1***	***F= 2***	***F = 3***
Employment	-0.019	0.094	0.114	0.135
	(0.115)	(0.100)	(0.146)	(0.160)
Wage	0.318***	0.310***	0.263*	0.141
	(0.116)	(0.119)	(0.170)	(0.170)
Productivity	0.138	-0.045	0.102	0.108
	(0.391)	(0.408)	(0.362)	(0.318)
Capital	0.048	0.174	0.198*	0.170
	(0.059)	(0.131)	(0.126)	(0.127)
Output	0.043	-0.030	0.103	0.207
	(0.339)	(0.371)	(0.354)	(0.336)
Skill intensity	----	----	----	----

*Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

It is also noticeable that employment positively increased in acquired firms when compared to relatively similar non-acquired firms. It is evident that, for acquired firms relative to similar but not acquired firms, the increase in employment after acquisition is above 10% on average. Similar to the main FE results, DID acquisition effect on employment was not statistically significant as the effects on wages. A visual display of the ATT results for some of the outcome variables is given in the plots in Figure 4.2. In each panel, the solid circles show the point estimates of ownership change (or acquisition), where it was assumed that the acquired firms remained foreign-owned throughout the study period. The vertical bars denote 95% confidence intervals based on 100 block bootstrap replicates. Referring to the plots, it is observed that the point estimates for the panel on wage are far from zero, especially from the acquisition year to the third year after acquisition. In addition, for capital, the distance between the point estimates and the horizontal dotted line widens from the acquisition year onwards. Regarding the rest of the outcome variables, these point estimates are close to zero, which indicates a picture reflective of the results in Table 4.4. It can also be noticed that generally variance tends to increase regarding each outcome variable irrespective of the nature of the effect. This is a probable indicator that firms tend to follow a heterogeneous path.

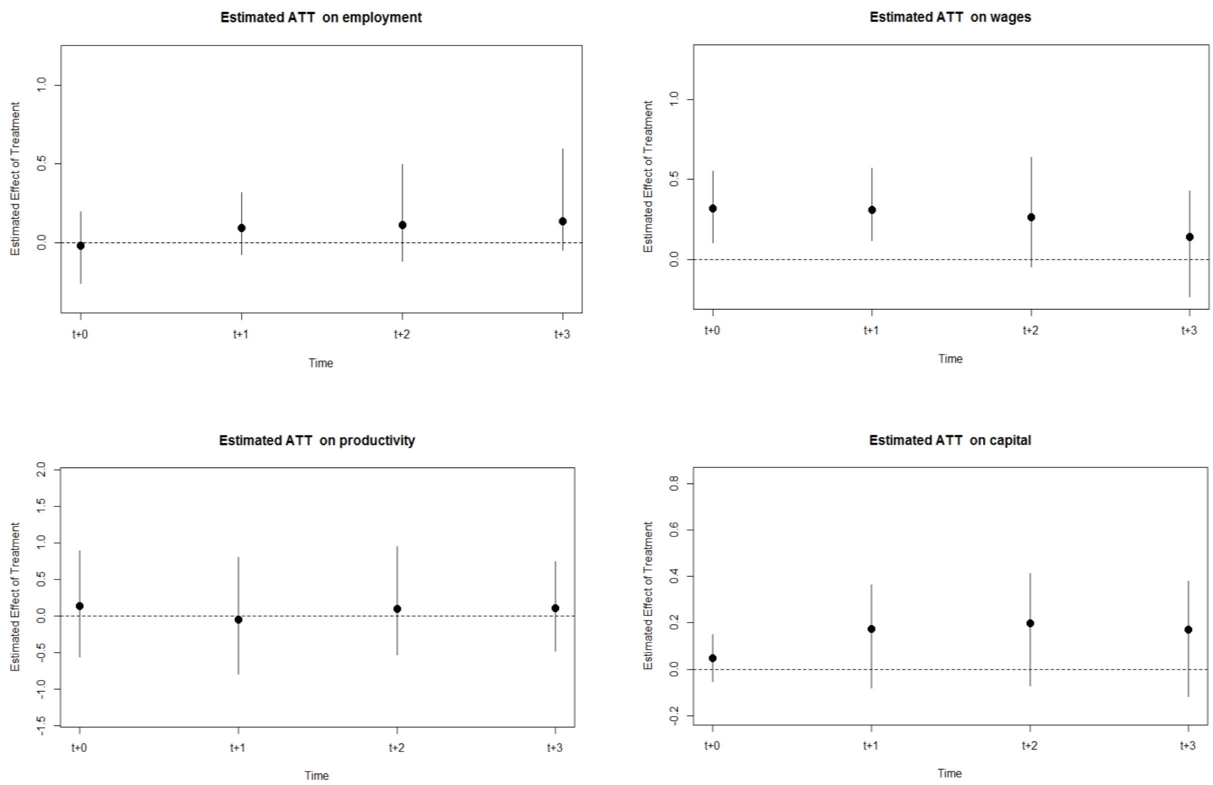


Figure 4. 2: Plots of difference-in-differences ATT for some of the outcomes under consideration

4.4.2.7 Covariate balance results

Results from matching analysis are merited when the baseline (pre-treatment) covariates used in the matching are balanced between the treated and untreated units. In this thesis' context, such a balance is what would merit this technique as a suitable robustness check for the main results obtained using the FE estimation methods. One of the most popular statistics used in assessing covariate balance in matching analysis, especially where propensity scores are used, is the standardised mean difference (SMD). This is because it is easy to compute and understand (Zhang, Kim, Lonjon & Zhu, 2019: 3). Utilising the specification in (4.11) and (4.12), I conducted balance checks for the covariates in the pre-acquisition period.

In Figure 4.3, plots of covariate balance due to matching over the pre-acquisition period are shown. For each outcome variable, the SMD is plotted on the vertical axis over the pre-acquisition period of two years on the horizontal axis. The panels in rows indicate the balance for outcome variables in the following order: employment, wage, productivity, output, capital and capital intensity. The columns correspond to the refinement methods used. Results indicate a good balance of majority variables on the two-year period. However, there are variations in the extent of the balance for specific outcome variables and with specific refinement methods. For instance, for employment as the outcome, propensity weighing (1:3) shows better balance when compared to the rest while, for wages, the propensity score refinement seems to provide better balancing on covariates. Yet on capital and capital intensity, Mahalanobis distance refinement seemingly provides better balance. Despite these slight variations, estimates of ATT based on the three refinement methods yield similar results, as will be seen later. It can also be noticed that before the two-period mark –this being the pre-treatment period – for panel (1:1) and (3:1), the employment and productivity outcome variables remain constant, although with some imbalance. This is a positive confirmation for the parallel trend assumption alluded to in (4.7), a fundamental assumption that authenticates my choice of the DID estimator for robustness checks on the main results. Other outcome variables, such as output and capital were not part of the probit model and so cannot be directly investigated regarding this assumption. Overall, the balance checks lent credence to the matching and DID estimation results presented and discussed earlier.

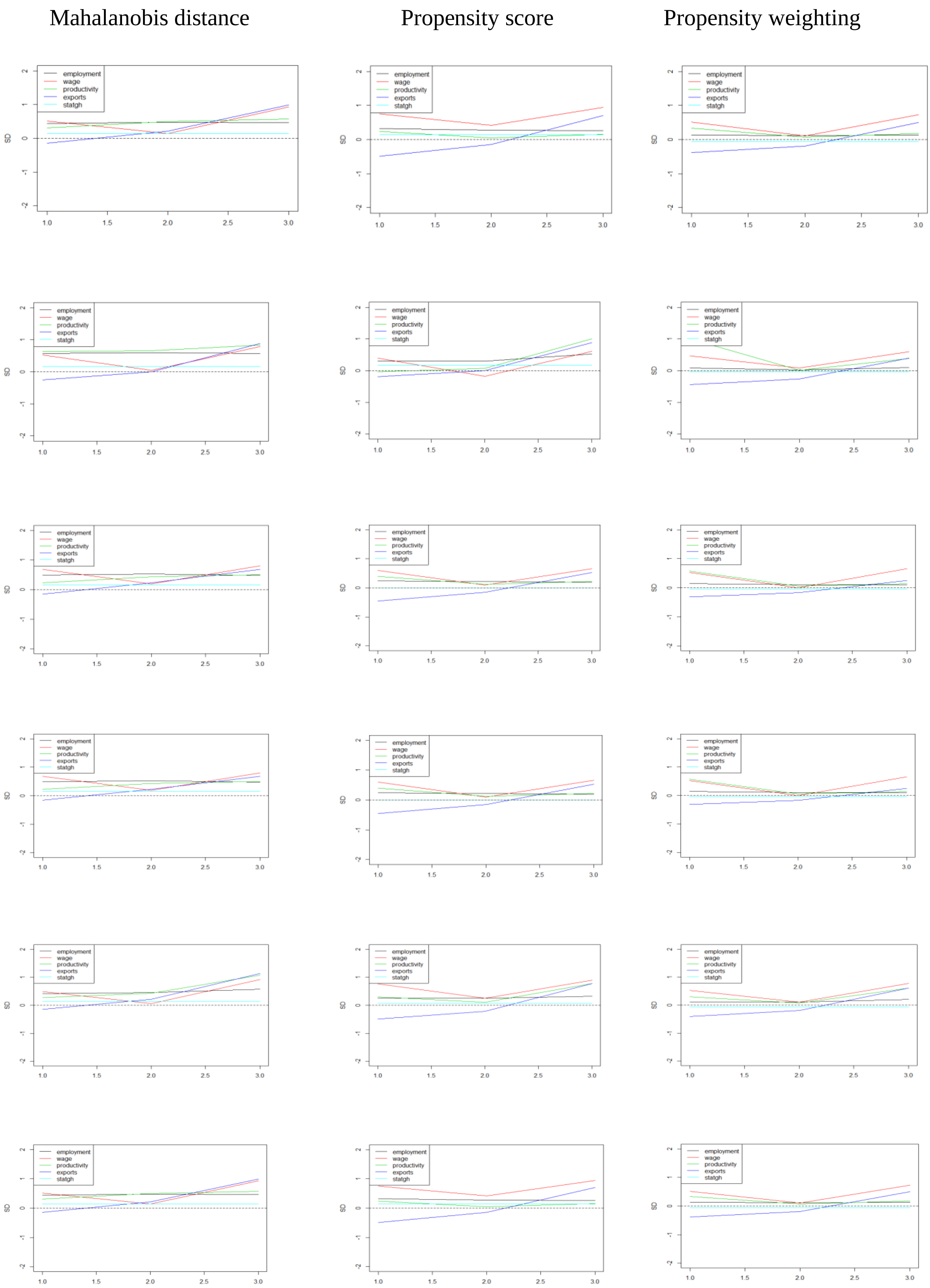


Figure 4. 3: Plots of covariate balance after matching

Besides meriting the matching results using covariate balance checks, I also performed further robustness checks specifically for the matching method by using other alternative matching and refinement methods. I specifically employed propensity scores and propensity weighting as refinement methods to re-estimate the average acquisition effect on the performance outcomes of acquired firms. The results from these two methods are shown in Table 4C in the appendix. These results are comparable in magnitude and statistical significance to those generated using Mahalanobis distance methods. This attests to the reliability of the preferred refinement method, Mahalanobis distance, used in the initial refinement of the matched sets.

4.5 Conclusion from findings in chapter four

A large body of research examines the link between foreign-owned investments and economic benefits that accrue to host economies. Investigation has been both at macro and micro levels, with tangible contributions pointed out over and above mere capital investments. As highlighted by Matthias and Javorcik (2009: 51), empirical focus has been guided largely by the assumption that foreign ownership solely is associated with tangible benefits both at macro level and at firm level. Whereas this assumption may hold for Greenfield investments, it is not clear in the case where foreign ownership arises out of acquisition of a formerly domestically-owned firm. This is due to the likelihood of selection bias, which, more often than not, drives the acquisition decision of most foreign investors. In this study, I focused on ownership arising out of acquisition and identified the effect while taking care of firm-level heterogeneity and endogeneity effects by applying regression methods and robustness checks using matching and difference-in-differences estimation methods.

This study differs from majority studies because, in addition to the usual outcomes, focus is extended to other outcomes thought to be influenced by foreign ownership. The study further looks at the effect on capital, one of the key impediments to investment and firm performance growth in developing countries. The country/region of focus is also worth noting as few causal studies on foreign ownership effects have been conducted in SSA.

The results from regressions generally reflect those from matching with few and negligible differences. Whereas regression methods handle firm-level heterogeneity, the method used for robustness checks attempts to handle probable endogeneity. The minor differences in the two estimation results may be reflective of sample size issues. Evidence from this chapter's findings indicates that foreign acquisition tends to cause significant rises in mean wages, productivity, and output and generally positive (but insignificant) effects on employment, capital and skill intensity levels in acquired firms. A key empirical conclusion from the results is that, as the motivational discussion highlighted, foreign ownership indeed has significant effects on firm level performance outcomes along variables like; productivity, output, and wages. This revelation is in agreement with numerous scholarly findings highlighted earlier. However, the fact that results also indicate insignificant acquisition effects on other variables like; capital, employment, and skill intensity, shows that there is truth in conclusions by scholars like Navaretti *et.al* (2004), and Weche Geluebck (2015). Despite the sample size issues noted in this analysis, it is evident that the link between foreign ownership and firm performance is not correlative as some reviewed literature asserts.

Currently, numerous developing economies, particularly in SSA, are in pursuit of policies aimed at attracting inward foreign owned investments. Such policies have been motivated by economic benefits believed to be associated with these investments, evidence of which has been mainly from the developed world and East Asian economies. By providing further evidence, this study cements the view that foreign acquisition (and consequent ownership) of domestic firms is, to a recognisable extent, responsible for improvements in firm performance. And by positively influencing firm-level performance, foreign investments impact on the overall wellness of host economies. Although this evidence is observed in Ghana, its applicability to SSA is undoubtable given the structural similarities amongst African economies and generally the developing world. Study findings in this chapter therefore lend credence to the current policy orientation in most SSA economies.

Chapter 5

Conclusions

5.1 Summary

International literature is characterised by evidence of performance superiority of foreign-owned investments over domestically-owned plants. This immediately elicits the question as to what accounts for the observed superior performance of these investments compared to local firms. Is it that these investments are systematically superior to (and hence different from) domestically-owned firms? Or do foreign investors simply decide to merge with or buy off those local firms that are already performing well? In the event that such foreign ownership indeed arises out of acquisition of a locally owned plant, to what extent should we attribute the resulting performance to foreign takeover? Do the stylised facts prevalent in international literature hold true for SSA economies where, despite increased inflows of foreign investments in recent years, economic outcomes such as exports and employment growth continue to be low? Answers to these and many other related empirical questions have long been the subject of scholarly investigation in the field of foreign direct investment literature. Thus far, available evidence remains inconclusive and biased in scope towards the Western world.

This thesis intended to provide additional empirical evidence particularly relating to SSA. In this endeavour, the thesis addressed two specific research questions. Firstly, do foreign-owned investments differ systematically from domestically-owned enterprises; or does comparing the two forms of investments equate to simply comparing sub-groups of firms (for instance, large firms, where foreign-owned investments are most likely to be found, versus small and medium-sized investments)? This empirical question was attended to in Chapters 2 and 3, which provided a country-specific perspective on Uganda, as well as a multi-country perspective focusing on 19 countries in SSA. In Chapter 4, the thesis addressed the second empirical question of whether foreign ownership (specifically arising out of acquisition of a formerly domestically-owned firm) has significant effects on firm-level performance.

One of the approaches to use in an attempt to answer these empirical questions is to examine firm-level characteristic and performance differences between foreign-owned and domestically-owned firms. This involves an examination of how the two forms of investments differ along particular performance dimensions at the micro-level. A second approach is to focus specifically on a single characteristic (being, in the case of this thesis, nature of ownership) and examine its likely effect on firm-level performance in the same way that Earle *et al.* (2012) have done. In an effort to provide the intended additional evidence on foreign-owned investments, this thesis adopted both approaches.

The evidence in this thesis came from three unique data sets, namely, two different firm panels and one comprehensive cross-sectional survey. The data sets had the unusual advantage of coming from SSA, a region hitherto neglected in earlier studies. Because of its geographical scope, this thesis therefore provides new and crucial evidence. Secondly, these data sets contain many foreign-owned firms (in the instance of the Uganda and UNIDO data) for better comparison and domestically-owned firms for control units (in the instance of the panel for Ghana).

The methods applied to these data included cluster analysis methods using unsupervised machine learning techniques, and regression analysis coupled with conventional impact evaluation methods. In Chapter 2, cluster analysis methods were applied. Specifically, using machine-learning techniques, I applied agglomerative clustering to generate homogeneous groups of firms based on carefully selected performance indicators. I applied descriptive statistical analysis on resultant clusters of firms. Estimated averages arising from the results revealed that foreign-owned investments were likely to be systematically different from domestically owned investments. The findings suggest that comparative differences between foreign-owned firms and domestically-owned firms are not indicative of different sub-groups of firms, but rather that foreign-owned firms are systematically different along numerous performance indicators. Indeed, in the classification and regression tree (CART) analysis, foreign ownership was found to be a key segmenting variable in the creation of clusters firms. In Chapter 3, similar methods of analysis were adopted to investigate whether findings in Chapter 2 hold equally for a multi-country case and for economic groupings of countries by internationally reputable institutions. The core findings concurred with those in Chapter 2, adding revelations of intra- foreign-owned firms' heterogeneities. Findings in both these chapters are also reflected in empirical findings in available international literature[71].

Regression methods and impact evaluation were used in Chapter 4 in order to investigate the effect of foreign acquisition on firm performance. In the context of acquisition or merging, one

[71] See studies by Njikam (2018), Coniglio *et al.* (2015), Amendolagine *et al.* (2013) and Athukorala *et al.* (1995).

of the prominent theoretical strands holds that the performance premium is a result of foreign-owned firms' transfer of technical and managerial skills to the acquired affiliate in the host economy. Any empirical test on this theoretical strand necessitates that the analyst rules out the likelihood of pre-acquisition selection bias when examining the different performances of acquired firms and those firms that remained purely domestic. Using the stated methods, I estimated fixed effects regressions for numerous firm-level outcomes. Results indicated positive and significant effects of foreign acquisition on wages (9.9% in favour of foreign firms), productivity (35.8%), and output (43.3%). Further positive though not significant effects were revealed for capital (17.4%), employment, and skill intensity. Because fixed effects regressions couldn't handle other econometric issues like endogeneity, I decided to employ impact evaluation methods as a robustness check for the regression results. I therefore matched firms on a set of pre-acquisition covariates, including lagged outcomes in order to construct a feasible control group of non-treated firms. I applied difference-in-differences estimation techniques and still found significant estimates of foreign acquisition effects on acquired firms lying in the positive range of 26% to 36% for wages; 5% to 20% on capital investments; and sizeable estimates (though not significant) for other outcomes, such as employment, output and productivity. While less significant than those generated by POLS and FE, these estimates, pointing to possible significant selection effect on the acquisition decision, are comparable to those from the FE and numerous empirical study findings in international literature[72]. These findings are also in disagreement with those of scholars such as Navaretti *et. al* (2004) and Weche Geluebcke (2015). Nevertheless, the estimates are highly suggestive that foreign ownership has positive effects on firm-level performance. Given that the empirical results from the analyses tend to support and reveal positive effects of foreign investment, yet the dismal economic outcomes observed in SSA and which motivated this investigation are observed at a time when inflows of such investments is on the rise, it becomes clear that these dismal outcomes are partly a result of other factors beyond the scope of this thesis' inquiry. This implies that although these findings may be used to guide policy, caution has to be observed especially regarding the limited scope of the inquiry.

This thesis contributes greatly to the growing empirical literature on foreign-owned investments. By using alternative methods in part of the analysis, it provides a robustness check

[72] See studies by; Wang &Wang (2015a), Conyon *et al.* (2002), Huttunen (2007), Fabling & Sanderson (2014), Matthias & Javorcik (2009), among others

of the findings of earlier studies. The thesis provides further empirical findings in support of the now stylized fact that foreign-owned firms differ from domestically-owned firms. Findings confirm that these differences are systematic along numerous performance characteristics. These differences hold not only for specific economies but also for multi-country scenarios in SSA. Through the application of conventional impact evaluation methods and regression in part of the analysis, the thesis further found compelling evidence that foreign ownership has positive effects on firm-level performance. Acquisition decisions of foreign investors are also found to be subject to selectivity that is driven by factors such as firm size. These findings are important in informing domestic economic policies in developing economies, particularly in SSA, where the urge to attract foreign direct investments is evident. Findings in this thesis seem to support the thinking that such investments can partly contribute to solving current socioeconomic problems such as unemployment, low export volumes, scarcity of investment capital, and general wellness of host economies.

A probable limitation of analyses in this thesis relates to data, especially in terms of the time when data collection was done. Analyses are based on data which is as recent as 2013. This is because these data are the most recent versions available for the specific countries selected for studies in this thesis. Findings in this thesis may thus be subject to recent changes in economic environments in the respective countries analysed and SSA in general. Although this is an insignificant limitation, caution should be observed when using the findings. Additionally, the geographical scope of Chapter 4 is also limited to only one country. It is possible that findings can vary across economies, even when similar phenomena are investigated using similar methods of analysis. Nevertheless, evidence from the three main chapters of this thesis is a noteworthy extension of scholarly literature on foreign direct investments, having important implications for policy as well. I briefly discuss some of these policy implications and potential areas for further scholarly research, in the next sections.

5.2 Policy Implications

From a general point of view, empirical evidence in this thesis shows that, at firm level, foreign-owned investments are on average likely to perform relatively better than domestically-owned investments in a typical host economy environment. If the view that performance of economies at macro level is partly reflective of the performance of production units such as firms at the micro level holds, then an economy with more foreign-owned investments would be expected to have better economic outcomes relative to one with fewer foreign-owned investments. Given this perspective, a probable policy suggestion to developing economies, such as those in SSA (where this thesis was focused), would be to increase their efforts towards attracting more foreign direct investments. Evidence of foreign acquisition effects on performance at firm level in Chapter 4 also supports this policy suggestion. By having direct effects on firm performance at the micro level, evidence in this thesis confirms that foreign direct investments, *ceteris paribus*, may be effective in contributing to the overall wellness of the host economy, and that any policy orientation towards attracting more foreign-owned investments would probably be a step in the right direction.

Findings in Chapter 4 suggest that foreign-owned firms most likely hire highly skilled workers; hence the high wage and productivity effects that are observed in foreign-owned firms. This might imply that if policymakers in a host economy decide to increase efforts to attract more foreign direct investments, more benefits are likely to be reaped if credible investments in human capital have been put in place. Based on this thesis' findings, this is because more skilled nationals are likely to be employed as foreign investment inflows increase. Moreover, the importance of human capital in attracting foreign direct investment has been empirically confirmed for both developed and developing economies. Specifically, the level of educational attainment of the host country has been revealed as important in the foreign direct investment context[73].

Estimation results using matching in Chapter 4 further suggest that firms tend to follow a heterogeneous growth path. A high likelihood of firm-level heterogeneities is further demonstrated in Chapter 3 among clusters with purely foreign-owned firms. This indicates that, even among foreign-owned firms, there are probable variations in firm-level characteristics such as productivity and size. These variations, however, usually lead to systematic variations in firm-level market behaviour, especially regarding firms' abilities to participate in international markets[74]. Reference to this thesis's contextual background reveals low exports amongst SSA economies as one of the key elements. Yet exports, especially those with high technology intensity have significant positive effects on economic performance (Wabiga & Nakijoba, 2018). Highly productive firms are more likely to export. It may then be credible that policies aimed at export stimulation through attracting and incentivising foreign direct investments need to take into account likely firm heterogeneities if one is to go by evidence

[73] See the study by Karimi, Yusop, Hook and Chin (2013) on the effect of human capital on FDI inflows.
[74] See Plouffe (2016) on Firm Heterogeneity and Trade-Policy Stances.

from this thesis. This will ensure that more ‘cherries’ (as opposed to ‘lemons’) are attracted and incentivised.

Finally, Chapter 2 contained the noteworthy finding that some foreign-owned firms may not significantly differ from domestically-owned firms in Uganda in terms of some performance indicators, such as export intensity. Indeed, clustering results indicated that a good number of foreign-owned firms is clustered in groups composed of mainly domestically-owned firms. This finding probably points to two policy-related issues: either the belief that numerous foreign investors target local markets in host economies as opposed to international markets, and may therefore not effectively contribute to the export desires of host economies; or that some domestically-owned investments actually match foreign-owned investments along some performance indicators. Some domestically-owned firms being clustered in groups largely composed of foreign-owned firms and posting higher mean values further corroborates the latter position. Given such findings, a probable implication could be that policies that seek to promote and hence optimise benefits from both domestic and foreign-owned investments, are relatively better than policies focused mostly on stimulating only one form of investment.

5.3 Suggestions for the Future

Studies such as that of Bentivogli *et al.* (2016), have found that the foreign ownership premium is only significant in the service sector. The justification is that such a sector is shielded from disciplinary measures in the international market, which gives the sector more room for performance enhancement. It would therefore be of great interest for researchers to extend my current analysis using data that covers both sectors. This would provide a robustness check for this empirical school of thought and provide better policy guidance.

One firm-level outcome not considered, especially in Chapter 4 of this thesis, is the export performance of acquired firms. Yet the effectiveness of foreign-owned investments in improving export performance of host economies has been supported by available literature. Chapters 2 and 3 further find exportation as a key firm-level characteristic that differentiates foreign-owned and domestic firms. Were data available, it would be relevant to test this hypothesis in the context of SSAs economies. This would provide evidence about whether foreign ownership positively affects the export abilities of firms, or whether foreigners simply acquire those local firms that are already export-oriented.

In order to deepen scholarly understanding of the probable welfare effects of foreign-owned investments on host economies from a microeconomic lens, future investigation may have to extend analysis to include more welfare dimensions, such as the quality of the work environment and labour market security. These two elements were not covered in this thesis due to data limitations. However, together with workers' earnings, they complete the set of elements used in measuring and assessing job quality, a key determinant of workers' welfare status. Some empirical studies have been conducted on whether foreign direct investments create quality jobs in host economies but this has mainly been done in Asian economies and the western world.

Finally, Chapter 4 of this thesis provides evidence on the effect of foreign ownership on firm performance, a topical issue that still elicits attention in most empirical investigations on foreign direct investments. This evidence, however, is provided on a specific country case. The current case of Ghana may distract us from the influences of critical drivers of firm performance growth, such as institutions. It would therefore be prudent for future analysis to perform the same from a multi-country perspective and to cater for important aspects, such as institutional quality.

References

Aaron, G., Iregui, A.M. & Ramírez, M.T. 2014. An Empirical Examination of the Determinants of Foreign Direct Investment : A Firm-Level Analysis na Colômbia : Evidencia ao nível de firma. *Revista de Economía del Rosario*. 17(1):5–31.

Abbas, O. 2008. Comparisons Between Data Clustering Algorithms. *Int. Arab J. Inf. Technol.* 5(3):320–325. [Online], Available: http://www.ccis2k.org/iajit/PDF/vol.5,no.3/15-191.pdf.

Ablov, A. 2015. The Firm-Level and Regional Determinants of FDI Distribution in Poland : Does Sector of Economy matter? *Economia XXI Wieku*. 4(8).

Adda, W.A. 1996. Privatisation in Ghana's Public Enterprise Reform Programme. *O. Fadahunsi (ed.), Privatization in Africa: The Way Forward. Nairobi: AAPAM*. 221–233. [Online], Available: http://unpan1.un.org/intradoc/groups/public/documents/aapam/unpan027477.pdf.

Aliber, R.Z. 1970. A Theory of Foreign Direct Investment'. *The International corporation*. 17–34.

Almeida, R. 2007. The Labor Market Effects of Foreign-owned Firms. *Journal of International Economics*. 72:75–96.

Anghel, B. 2007. A Knowledge-Capital Model Approach of FDI in Transition Countries. *Universitat Autonoma de Barcelona*. (November 2006):1–35.

Aydin, N., Sayim, M. & Yalaman, A. 2007. Foreign Ownership and Firm Performance : Evidence from Turkey. *International Research Journal of Finance and Economics ·*. 11(January):103–111.

Baltagi, B.H. 2005. *Econometric Analysis of Panel Data*. Third Edit ed. West Sussex: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,West Sussex PO19 8SQ, England Telephone.

Bandick, R. 2011. Foreign acquisition, wages and productivity. *The World Economy*. 34(6):931–951.

Banerjee, A. & Davé, R.N. 2004. Validating clusters using the Hopkins statistic. *IEEE International Conference on Fuzzy Systems*. 1:149–153.

Barba Navaretti, G., Checchi, D. & Turrini, A. 2002. Adjusting Labour Demand: Multinational vs. National Firms. A Cross-European Analysis. *SSRN Electronic Journal*. (August):1–21.

Barrell, R. & Pain, N. 1996. An Econometric Analysis of U . S . Foreign Direct Investment. *The Review of Economics and Statistics*. 78(2):200–207. [Online], Available: https://www.jstor.org/stable/2109921.

Barthel, F., Busse, M. & Osei, R. 2008. The Characteristics and Determinants of FDI in Ghana. *The European Journal of Development Research*. 23(3):389–408.

Basu, A. & Krishna Srinivasan. 2002. *Foreign Direct Investment in Africa: Some Case Studies*. (WP/02/61). Washington DC.

Beckstead, J.W. 2002. Using Hierachical Cluster Analysis in Nursing Research. *Western Journal of Nursing Research*. 24(3):307–319.

Benfratello, L. & Sembenelli, A. 2006. Foreign ownership and productivity : Is the direction of causality so obvious ? *International Journal of Industrial Organization*. 24:733–751.

Bentivogli, C. & Mirenda, L. 2016. Foreign ownership and performance : Evidence from a panel of Italian firms. *International Journal of the Economics of Business*. 24(September):251–273.

Bishop, K., Mason, G. & Robinson, C. 2009. *Firm growth and its effects on economic and social outcomes*.

Blanas, S., Seric, A. & Viegelahn, C. 2017. *Jobs , FDI and institutions in Sub-Saharan Africa : Evidence from firm-level data*. Geneva.

Blashfield, R.K. 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*. 83(3):377–388.

Blomstrom, M. & Lipsey, R.E. 1986. *Firm Size and Foreign Direct Investment*. Cambridge.

Blonigen, B.A., Davies, R.B. & Head, K. 2003. Estimating the Knowledge-Capital Model of the Multinational Enterprise : Comment. *American Economic Review*. 93(3):980–994.

Bock, H. 2008. Origins and extensions of the k -means algorithm in cluster analysis. *Journal Electronique d'Histoire des Probabilités et de la Statistique*. 4(December):1–18.

Bouras, H. & Raggad, B. 2015. Foreign Direct Investment and Exports : Complementarity or

Substitutability An Empirical Investigation. *International Journal of Economics and Financial Issues*. 5(4):933–941.

Bprgen, F.H. & Barnett, D.C. 1987. Applying Cluster Analysis in Counseling Psychology Research. *Journal of Counceling Psychology*. 34(4):456–468.

Braconier, H., Norbäck, P. & Urban, D. 2002. *Vertical FDI Revisited*. (167).

Brainard, S.L. 1993a. *A simple theory of Multinational Corporations and Trade with a Trade-off between proximity and concentration*. (4269). Cambridge.

Brainard, S.L. 1993b. *An empirical assessment of the proximity concentration tradeoff between multinational sales and trade*. (4580). Cambridge.

Brambilla, I., Lederman, D. & Porto, G. 2012. Exports, Export Destinations, and Skills. *American Economic Review*. 102(7):3406–3438.

Brauksa, I. 2013. Use of Cluster Analysis in Exploring Economic Indicator Differences among Regions : The Case of Latvia. *Journal of Economics, Business and Management*. 1(1):42–45.

Bricongne, J.-C., Bedoya, S.F. & Forero, M.L. 2016. The Proximity-Concentration Trade-off with Multi-product Firms : Are Exports and FDI Complements or Substitutes? (February):0–36.

Buckley, P.J. & Casson, M.C. 2009. The internalisation theory of the multinational enterprise: A review of the progress of a research agenda after 30 years. *Journal of International Business Studies*. 40(9):1563–1580.

Cai, H. & Guney, Y. 2010. European Union Foreign Direct Investment in China: Evidence from a Panel Study of EU Manufacturing Firms, 1998-2007. In Cambridge *15th Cambridge International Manufacturing Symposium, 23rd?24th September 2010, Cambridge, UK*. [Online], Available: http://eprints.hud.ac.uk/id/eprint/17214/.

Caliñski, T. & Harabasz, J. 1974. A Dendrite Method For Cluster Analysis. *Communications in Statistics*. 3(1):1–27.

Can, M., DOĞAN, B. & DEĞER, O. 2017. The Relationship between Research & Development Investment Expenditure, Foreign Direct Investment and Economic Growth: Panel Causality and Cointegration Analysis for G-7 Countries. *Journal of*

Applied Economic Sciences. XII(1(47)):58–69.

Carr, D.L., Markusen, J.R. & Maskus, K.E. 2003. Estimating the Knowledge-Capital Model of the Multinational Enterprise : Reply. *American Economic Review*. 91(3):693–708.

Caves, R.E. 1971. The Industrial Economics of Foreign Investment. *Economica New Series*. 38(149):1–27.

Chari, A., Chen, W. & Dominguez, K. 2009. Foreign Ownership and Firm Performance : Emerging Market Acquisitions in the United States. *IMF Economic Review*. 60(590):1–42.

Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. 2014. NbClust : An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*. 61(6):1–36.

Chavent, M., Kuentz, V., Liquet, B. & Saracco, L. 2011. ClustOfVar: An R Package for the Clustering of Variables. *Journal of Statistical Software*. VV(II).

Chen, G., Geiger, M. & Fu, M. 2015. *Manufacturing fDi in Sub-Saharan africa : Trends, Determinants, and Impact*. Washington DC.

Christos, L., Konstantinos, V., Alexandros, G. & Xanthi, P. 2016. Manufacturing Firms ' Performance and Productivity : Evidence from North and South European , Scandinavian and Balkan Countries. *Theoretical Economics Letters*. 6(August):789–797. [Online], Available: http://dx.doi.org/10.4236/tel.2016.64083.

Cipollina, M., Pozzolo, A.F., Giovannetti, G. & Filomena, P. 2012. FDI and growth : what cross-country industry data say. *The World Economy*. 35(11):1599–1629.

Clinton, D., Button, E., Norring, C. & Palmer, R. 2004. Cluster analysis of key diagnostic variables from two independent samples of eating-disorder patients : evidence for a consistent pattern. *Journal of Psychological Medicine*. 34(May):1035–1045.

Coase, R.H. 1937. The Nature of the Firm. *Economica n.s*. 386–405.

Coniglio, N.D., Prota, F. & Seric, A. 2015. Foreign Direct Investment, Employment and Wages in Sub- Saharan Africa. *Journal of International Development*. 27(August):1243–1266.

Conyon, M., Girma, S., Thompson, S. & Wright, P. 2002. The impact of foreign acquisition on wages and productivity in the United Kingdom. *Journal of Industrial Economics*. 50(1):85–102.

Culem, C.G. 1988. The locational Determinants of Direct Investments among Industrialized Countries. *European Economic Review*. 32:885–904.

Davies, S.W. & Lyons, R. 1991 Characterising relative performance: the productivity advantage of foreign-owned firms in the UK

Delgado, M., Porter, M.E. & Stern, S. 2014. Clusters , convergence , and economic performance. *Research Policy*. 43(10):1785–1799.

Demurger,C. &Fournier, M. 2005 Earnings differentials and ownership structure in Chinese enterprises. *Journal of economic development and cultural change*.53(4): 933-958

Denisia, V. 1998. Foreign Direct Investment Theories: An Overview of the Main FDI Theories. *European Journal of Interdisciplinary Studies*. 2(2):53–59.

Dias, A., Pinto, C. & Batista, J. 2016. *Signaling Tax Evasion, Financial Ratios and Cluster Analysis*. (51). [Online], Available: http://www.gestaodefraude.eu.

Díaz-bonilla, E. & Thomas, M. 2016. Why Some Are More Equal Than Others: Country Typologies of Food Security. *International Food Policy Research Institute*. 1510(February).

Driver, H.E. & Kroeber, A.L. 1932. Quantitative Expressions of Cultural Relationships. *University of California*. 31(4):211–256.

Dunning, J.H. 1977. *Trade, location of economic activity and the multinational enterprise: A search for an eclectic approach*. Springer.

Dunning, J.H. 1979. Explaining changing patterns of international production: in defence of the eclectic theory. *Oxford bulletin of economics and statistics*. 41(4):269–295.

Dunning, J.H. 1980. Toward an Eclectic Theory of International Production: Some Empirical Tests. *Journal of International Business Studies*. 11(1):9–31.

Dunning, J.H. 1988. The Theory of International Production. *International Trade Journal*. 3(1):21–66.

Dunning, J.H. 1993. Multinational Enterprises and the Global Economy. *Mass: Addison*

Wesley.

Dunning, J.H. 2003. The eclectic (OLI) paradigm of international production: Past, present and future. *International Business and the Eclectic Paradigm: Developing the OLI Framework*. 8(2):21–39.

Earle, J.S., Telegdy, A. & Antal, G. 2012. FDI and Wages : Evidence from Firm-Level and Linked Employer-Employee Data in Hungary , 1986-2008. *IZA Discussion Paper No. 7095*. (7095):1986–2008.

Edward, J.R. 1989. *The Determinants of Foreign Direct Investment in the United States, 1979 - 85*. University of Chicago Press. [Online], Available: http://www.nber.org/chapters/c6176.

Erdal, L. & Göçer, İ. 2015a. The Effects of Foreign Direct Investment on R&D and Innovations: Panel Data Analysis for Developing Asian Countries. *Journal-Procedia - Social and Behavioral Sciences*. 195:749–758.

Erdal, L. & Göçer, İ. 2015b. The Effects of Foreign Direct Investment on R&D and Innovations: Panel Data Analysis for Developing Asian Countries. *Jornal -Procedia - Social and Behavioral Sciences*. 195(July):749–758.

Everitt, B.S., Landau, S., Leese, M. & Stahl, D. 2011. *Cluster analysis*. 5th Editio ed. S.W. David J. Balding, Noel A.C. Cressie, Garrett M. Fitzmaurice, Harvey Goldstein, Geert Molenberghs, David W. Scott, Adrian F.M. Smith, Ruey S. Tsay (ed.). John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom For. [Online], Available: http://www.wiley.com/ WileyCDA/Section/id-300611.html.

Fabling, R. & Sanderson, L. 2014. Foreign acquisition and the performance of New Zealand firms. *New Zealand Economic Papers*. 48(1):1–20.

Faeth, I. 2009. Determinants of foreign direct investment - A tale of nine theoretical models. *Journal of Economic Surveys*. 23(1):165–196.

Falk, M. & Wolfmayr, Y. 2010. The Extent, Characteristics and Impacts of FDI and Multinational Firm Activities - A Firm Level Analysis. *FIW Research Reports*. (November 2008). [Online], Available: http://hdl.handle.net/10419/121212.

Fânaru, M. 2016. Multivariate analysis of marketing data - applications for bricolage market.

Bulletin of the Transilvania University of Brasov. Economic Sciences. 9(2):419.

Feng, Y. 2017. Determinants of Foreign Direct Investment (FDI). *Oxford Research Encyclopedia of Politics*. (August):1–18.

Ferreira, L. & Hitchcock, D.B. 2009. A Comparison of Hierarchical Methods for Clustering Functional Data. *Communications in Statistics-Simulation and Computation*. 38(9):1925–1949.

Forssbeck, J. & Oxelheim, L. 2008. Finance-specific factors as drivers of cross-border investment—An empirical investigation. *International Business Review*. 17(6):630–641.

Forte, R. & Santos, N. 2015. A cluster analysis of FDI in Latin America. *Latin American Journal of Economics*. 52(1):25–56.

Foster-McGregor, N., Isaksson, A. & Kaulich, F. 2013. *Importing, Exporting and Performance in Sub-Saharan African Manufacturing Firms*. (96). Vienna.

Gallegos, M.T. & Ritter, G. 2005. A robust Method for Cluster Analysis. *The Annals of Statistics*. 33(1):347–380.

Geho, L. and L. 2018. LafargeHolcim Annual Report 2017. Building for Growth.

Georgopoulos, A. & Preusse, H.G. 2009. Cross-border acquisitions vs. Greenfield investment: A comparative performance analysis in Greece. *International Business Review*. 18(6):592–605.

Gerlach, A.-C. & Liu, P. 2010. Resource-seeking Foreign Direct Investment in African Agriculture A review of country case studies. (31):1–19. [Online], Available: http://www.fao.org/fileadmin/templates/est/PUBLICATIONS/Comm_Working_Papers/EST-WP31.pdf.

Girma, S. & Görg, H. 2007a. Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics*. 72:97–112.

Girma, S. & Görg, H. 2007b. Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics*. 71(3):97–112.

Giulietti, M., Mccorriston, S. & Osborne, P. 2004. Foreign direct investment in the UK :

evidence from a disaggregated panel of the UK food sector. *Journal of Applied Economics*. 36(7):653–663.

Görg, H., Strobl, E. & Walsh, F. 2007. Why do foreign-owned firms pay more? the role of on-the-job training. *Review of World Economics*. 143(3):464–482.

Goswami, C. & Kanta, K. 2012. FDI and its relation with exports in India , status and prospect in north east region. *Procedia-Social and Behavioral Sciences*. 37:123–132.

Gower, J.C. & Legendre, P. 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*. 3(1):5–48.

Greenaway, D. & Kneller, R. 2007. Firm heterogeneity, exporting and foreign direct investment. *The Economic Journal*. 117(1996):F134–F161.

Grein, A.F., Sethi, S.P. & Tatum, L.G. 2010. A Dynamic Analysis of Country Clusters , the Role of Corruption , and Implications for Global Firms. *East-West Journal of Economics and Business*. XII(2):6–7.

Guadalupe, B.M., Kuzmina, O. & Thomas, C. 2012. Innovation and Foreign Ownership. *American Economic Review*. 102(7):3594–3627.

Gutiérrez-portilla, P., Maza, A. & Villaverde, J. 2016. Foreign direct investment in the Spanish regions : What are the influencing factors ? *Journal of Regional Research*. 35(35):67–82.

Hagemejer, J. & Tyrowicz, J. 2012. Is the effect really so large? Firm-level evidence on the role of FDI in a transition economy. *Economics of Transition*. 20(2):195–233.

Hale, G. & Mingzhi Xu. 2016. *FDI effects on the labor market of host countries*. (Working Paper 2016-25). San Francisco. [Online], Available: http://www.frbsf.org/economic-research/publications/working-papers/wp2016-25.pdf%0AThe.

Hamermesh, D.S. 1988. *Labor Demand and the Structure of Adjustment Costs*. (2572). Cambridge.

Hanafy, S. & Marktanner, M. 2019. Sectoral FDI, absorptive capacity and economic growth–empirical evidence from Egyptian governorates. *Journal of International Trade and Economic Development*. 28(1):57–81.

Hands, S. & Everitt, B. 1987. A Monte Carlo study of the recovery of cluster structure in binary data by hierachical clustering techniques. *Multivariate Behavioral Research*. 22(2):235–243.

Helpman, E. 1984. A Simple Theory of International Trade with Multinational Corporations. *Journal of Political Economy*. 92(3):451–471.

Henley, J., Kratzsch, S., Külür, M. & Tandogan, T. 2008. Foreign Direct Investment from China , India and South Africa in sub-Saharan Africa : A New or Old Phenomenon ? *World Institute for Development Economics Research*. 24.

Hennart, J.-F. 1982. A Theory of Multinational Enterprise . *University of Michigan Press*.

Hennart, J.-F. 1991. The transaction cost theory of the multinational enterprise. *The nature of the transnational firm*. 2:81–116.

Hilber, C.A.L. & Voicu, I. 2010. Agglomeration Economies and the Location of Foreign Direct Investment : Empirical Evidence from Romania. *Regional Studies*. 44(3):355–371.

Hollenstein, H. 2003. Innovation modes in the Swiss service sector : a cluster analysis based on firm-level data. 32:845–863.

Horst, T. 1972. Firms and industry determinants of the decision to invest abroad: an empirical study. *The review of economics and statistics*. 54(3):258–266.

Horstmann, J. & Markusen, J.R. 1996. Endogenous international market structures in trade (natura facit saltum). *Journal of International Economics*. 32(1992):109–129.

Hosseini, H. 2005. An economic theory of FDI: A behavioral economics and historical approach. *Journal of Socio-Economics*. 34(4):528–541.

Huttunen, K. 2007. The effect of foreign acquisition on employment and wages: Evidence from finnish establishments. *Review of Economics and Statistics*. 89(3):497–509.

Hymer, S. 1960. On multinational corporations and foreign direct investment.

Hymer, S.H. 1976. The International Operations of National Firms: A study of direct foreign investiment. *MIT Press*.

Illovo Sugar ltd. 2002. *Illovo Sugar; Annual Integrated Report*.

Imai, K., Wang, E. & Kim, I.S. 2019. Matching Methods for Causal Inference with Time

Series Cross Sectional Data. In Center for the Study of American Politics | Yale University — ISPS *CSAP Quantitative Research Methods Workshop*.

Imbens, G.W. 2015. *Matching Methods in Practice: Three Examples*.

Imbens & Wooldridge, J.M. 2007. Estimation of Average Treatment Effects Under Unconfoundedness. *Journal of Experimental Psychology: General*. 136(1):23–42.

Insights, P. 2018. Cement ' s Changing Landscape. (June).

International Labor Organization. 2019. *World Employment Social Outlook*.

International Monetary Fund. 2018. *World Economic and Financial Surveys- Regional Economic Outlook- Sub-Saharan Africa*. Washington DC.

Islam, M.A., Alizadeh, B.Z., van den Heuvel, E.R., Bruggeman, R., Cahn, W., de Haan, L., Kahn, R.S., Meijer, C., et al. 2015. A comparison of indices for identifying the number of clusters in hierarchical clustering: A study on cognition in schizophrenia patients. *Communications in Statistics: Case Studies, Data Analysis and Applications*. 1(2):98–113.

J.C.Gower. 1971. A General Coefficient of Similarity and Some of its Properties. *International Biometric Society*. 27(4):857–871.

Javorcik, B.S. 2015. Does FDI Bring Good Jobs to Host Countries? *The World Bank Research Observer*. 30(1):74–94.

Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*.

de Jonge, E. & van der Loo, M. 2013. An introduction to data cleaning with R. *Statistics Netherlands*. 53.

Kamata, I., Sato, H. & Tanaka, K. 2017. The Internationalisation of Firms and Management Practices : A Survey of Firms in Viet Nam. *ERIA Discussion Paper Series*. DP-2016-34(34):1–42.

Kaplinsky, R. & Morris, M. 2009. Chinese FDI in su Saharan Africa : Egaging with Large Dragons. *European Journal of Development Research*. 21(4):551–569.

Karlsson, S., Lundin, N., Sjöholm, F. & He, P. 2009. Foreign firms and Chinese employment. *World Economy*. 32(1):178–201.

Karpaty, P. & Andreas, P. 2006. The determinants of FDI flows : Evidence from Swedish manufacturing and service sector. *The Swedish Network for European Studies in Economics and Business*. (May).

Kassambara, A. 2017. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. Vol. 1. STHDA (http://www.sthda.com).

Kimura, F. & Kiyota, K. 2006. Exports , FDI , and Productivity : Dynamic Evidence from Japanese Firms. *Review of World Economics*. 142(4):695–719.

Kindleberger, C. 1969. American business abroad- Six lectures on direct investment. *New Haven and London: Yale University Press*.

Knickerbocker, F.T. 1973. Oligopolistic reaction and multinational enterprise. *The International Executive*. 15(2):7–9. [Online], Available: http://linkinghub.elsevier.com/retrieve/pii/0022190273802738.

Konings, J. & Vanormelingen, S. 2015. The Impact of Training on Productivity and Wages : Firm Level Evidence. *Review of Economics and Statistics*. 97(2):485–497.

Krugman, B.P. 1983. New Theories of Trade Among Industrial Countries. *The American Economic Review*. 73(2):343–347. [Online], Available: https://www.jstor.org/stable/1816867.

Krugman, P. 1991. Increasing Returns and Economic Geography. *Journal of Political Economy*. 99(3):483–499.

Kuah, A.T.H. 2002. Cluster Theory and Practice : Advantages for the Small Business Locating in a Vibrant Cluster. *Journal of research in marketing and entrepreneurship*. 4(3):206–228.

Lall, S. 1980. *The Pattern of Intra-Firm Exports by US Multinationals. In: The Multinational Corporation*. London: Springer.

Lazzeretti, L., Sedita, S.. & Caloffi, A. 2012. The birth and the rise of the cluster concept. *Paper presented at the DRUID*. 21.

Lewis, R.J. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. California.

Liao, M., Li, Y., Kianifard, F., Obi, E. & Arcona, S. 2016. Cluster analysis and its application to healthcare claims data : a study of end-stage renal disease patients who initiated hemodialysis. *BMC Nephrology*. 1–14.

Libanda, J., Nyasa, L. & Marshall, D. 2017. The Effect of Foreign Direct Investment on Economic Growth of Developing Countries : The Case of Zambia. *British Journal of Economics, Management & Trade*. 16(J2):1–15.

Lin, F. 2010. The determinants of foreign direct investment in China : The case of Taiwanese fi rms in the IT industry. *Journal of Business Research*. 63(5):479–485.

Lipsey, R.E. 2008. Foreign Takeovers and Employment in Indonesian Manufacturing. *Journal of Economic Literature*. 1–32.

Liu, W.-H. & Nunnenkamp, P. 2011. Domestic repercussions of different types of FDI:Firm-level evidence for Taiwanese manufacturing. *World Development*. 39(5):808–823.

Lombardo, R. & Falcone, M. 2011. Crime and Economic Performance. A Cluster Analysis of Panel Data on Italy'S Nuts 3 Regions. *Universit della Calabria, Dipartimento di Economia, Statistica e Finanza*. 0–33. [Online], Available: http://www.ecostat.unical.it/.

Love, J.H. & Lage-hidalgo, F. 2010. Analysing the determinants of US direct investment in Mexico. *Journal of Applied Economics*. 32(10):1259–1267.

Lynn, M. 2011. Segmenting and Targeting Your Market : Strategies and Limitations.

MacDougall, G.D.A. 1960. The benefits and costs of private investment from abroad: a theoretical approach. *Bulletin of the Oxford University Institute of Economics and Statistics*. 22(3):189–211.

MacManus, J.. 1972. The theory of the international firms. *G.Pacquet (ed), The Multinational Firm and the Nation State. Toronto: Collier-Macmillan*.

Maioli, S., Gong, Y. & Yundan, G. 2007. Employment Effects of Privatisation and Foreign Acquisition of Chinese State-Owned Enterprises. *International Journal of the economics of business*. 14(2453):197–214.

Makoni, P.L. 2015. An extensive exploration of theories of foreign direct investment. *Risk governance & control: financial markets & institutions*. 5(2):77–83.

Mangaraj, S. & Senauer, B. 2001. *A Segmentation Analysis of US Grocery Store Shoppers*. (1–8). Minnesota.

Manpreet kaur & Usvir Kaur. 2013. Comparison Between K-Mean and Hierarchical Algorithm Using Query Redirection. *International Journal of Advanced Research in Computer Science and Software Engineering*. 3(7):1454–1459.

Mariel, P., Orbe, S. & Rodr, C. 2007. The Knowledge-Capital Model of FDI : A time varying coefficients approach. *Scottish Journal of Political Economy*. 56(2):196–212.

Markusen, J.. & Venables, A.. 2000. The theory of endowment, intra-industry and multi-national trade. *Journal of international economics*. 52(2):209–234.

Martins, E.C. da S., Dias, J.A.S. & Triches, D. 2017. The determinants of foreign direct investment in Brazil : empirical analysis for 2001-2013. *CEPAL Review*. 172(121):171–184.

Mathä, T. 1999. Proximity-Concentration vs . Factor Proportions Explanation : The Case of Swedish Multinationals in the EU. *The European Institute of Japanese Studies*.

Matthias, J. & Javorcik, B.S. 2009. Gifted kids or pushy parents ? Foreign direct investment and plant productivity in Indonesia. *Journal of International Economics*. 79(1):42–53.

McManus, P.A. 2011. Panel Data Analysis October 2011. In Indiana: Indiana University *Introduction to Regression Models for Panel Data Analysis*. 1.

Melitz, M.J. 2002. *The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity*. (8881). Washington. [Online], Available: http://www.nber.org/papers/w8881.

Milligan, G.. & Cooper, M.. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*. 50(2):159–179.

Minas, C., Jacobson, D. & Mcmullan, C. 2014. Welfare regime , welfare pillar and southern Europe. *Journal of European Social Policy*. 24(2):135–149.

Mooi, E. & Marko, S. 2011. A concise guide to market research : The process , data , and methods using IBM SPSS Statistics . *Springer*. (January):18–19.

Moran, T.H., Alfaro, L. & Javorcik, B.S. 2007. How to Investigate the Impact of Foreign Direct Investment on Development and Use the Results to Guide Policy. *Brookings Trade Forum*. (2007):1–60. [Online], Available: https://www.jstor.org/stable/25063217.

Moss, T. & Ramachandran, V. 2004. Foreign Investment and Economic Development :

Evidence from Private Firms in East Africa. (41).

Mukiibi, M. & Bukenya, J.O. 2008. Segmentation Analysis of Grocery Shoppers in Alabama. In *Southern Agricultural Economics Association Annual Meeting, Dallas, TX*.

Mummolo, J. & Peterson, E. 2017. Improving the Interpretation of Fixed Effects Regression Results. *Journal of Political Science Research and Methods*. 6(issue 4):829–835.

Murry, M.J. 2016. Factor analysis, cluster analysis, and nonparametric research methods for heterodox economic analysis. In Cheltenham-UK: Edward Elgar Publishing *Handbook of Research Methods and Applications in Heterodox Economics*. 190–209.

Namisango, E., Harding, R., Katabira, E.T., Siegert, R.J., Powell, A., Atuhaire, L., Moens, K., Taylor, S., et al. 2015. A novel symptom cluster analysis among ambulatory HIV / AIDS patients in Uganda. *AIDS Care*. 27(8):954–963.

Naughtin, T.L. & Rankin, N. 2016. Firm Productivity , International Trade and Competition : Using micro data to examine the dynamics of South African firms. Universiteit Stellenbosch.

Navaretti, G.B., Cooper, R.N. & Venables, A.J. 2004. *Multinational Firms in the World Economy*. Princeton, NJ, Princeton University Press, 2004.

Neumann, L. & Kumwenda, H. 2017. *Malawi Extractive Industries Transparency Initiative (MWEITI): Scoping Study (Minerals , Oil and Gas , Forestry)*.

Nikola, Š. 2015. The Hierarchical Clustering of Tax Burden in the EU27. *Journal of Competitiveness*. 7(3):95–109.

Niringiye, A. & Tuyiragize, R. 2010. Determinants of a Firm's Level of Exports : Evidence from Manufacturing Firms in Uganda. *African Economic Research Consortium*. (May):1–40.

Njikam, O. 2018. Export market destination and performance : Firm-level evidence from Sub-Saharan Africa ☆. *Journal of African Trade*. 4(1–2):1–19.

O'Brien, R. & Williams, M. 2013. *Global Political Economy: Evolution and Dynamics*. second ed. New York: Macmillan.

Obwona, M.B. 1998. *Determinants of Foreign Direct Investments and their Impact on Economic Growth in Uganda*. (113761). Kampala.

OECD. (2008) Annual Reports

Orazem, P.F. & Vodopivec, M. 2004. Do Market Pressures Induce Economic Efficiency? The Case of Slovenian Manufacturing, 1994-2001 Peter F. Orazem. *Southern Economic Journal*. 76(3189):553–576.

Pacáková, Z. & Poláčková, J. 2013. Hierarchical Cluster Analysis – Various Approaches to Data Preparation. *AGRIS on-line Papers in Economics and Informatics*. 5(3):53–63.

Peluffo, A. 2015. Foreign Direct Investment , Productivity , Demand for Skilled Labour and Wage Inequality : An Analysis of Uruguay. *The World Economy*. 38(6):962–983.

Pfeiffer, B., Görg, H. & Perez-Villar, L. 2014. The Heterogeneity of FDI in Sub-Saharan Africa – How Do the Horizontal Productivity Effects of Emerging Investors Differ from Those of Traditional Players? *SSRN 2555547*. (262).

Phiri, W. 2011. Foreign Direct Investment in Zambia's Mining and Other Sectors: Distinguishing Drivers and Implications for Diversification. *Macroeconomic and Financial Management Institute of Eastern and Southern Africa*. 77–80.

Potter, A. 2015. Privatisation in Ghana Successes During Economic Collapse and Authoritarianism. *CUTS Centre for Competition, Investment and Economic Regulation (CUTSCCIER)*. (September):1–15. [Online], Available: http://www.cuts-ccier.org/pdf/Privatisation_in_Ghana-Alan_Potter.pdf.

Puron-cid, J.R.G.G. 2014. Using Panel Data Techniques for Social Science Research : an Illustrative Case and Some Guidelines. *CIENCIA ergo-sum*. 21:203–216.

Raff, H. & Ryan, M.J. 2008. Firm-Specific Characteristics and the Timing of Foreign Direct Investment Projects. *Review of World Economics*. 144(1):1–31.

Raff, H., Michael, R. & Staehler, F. 2006. *Asset Ownership and Foreign-Market Entry*. (1676).

Rankin, N. 2013. Exporting and Export Dynamics Among South African Firms. *South African Institute of International Affairs*. 149(June):1–22.

Rankin, N., Söderbom, M. & Teal, F. 2005. *Exporting from manufacturing firms in Sub-Saharan Africa E*. (GPRG_WPS-036). [Online], Available: http://www.gprg.org/.

Řezanková, H. 2014. Cluster Analysis of Economic Data. *Statistika: Statistics and Economy*

Journal. 94(1):73–86.

Riddervold, S. & Kristiansen, S. 2011. The Effects of Foreign Direct Investment on the Ugandan Economy. Universitetet i Agder.

Rodríguez, A. & Tello, P. 2014. Foreign Direct Investment Impact on the Productivity and Employment of Spanish Manufacturing Firms (2001-2010) 1. *Journal of Economic Literature*. (November):1–23.

Rousseeuw, P.J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 20(4):53–65.

Rovan, J. & Sambt, J. 2003. Socio-economic Differences Among Slovenian Municipalities : A Cluster Analysis Approach. *Development in Applied Statistics*. 265–278.

Rugman, A. 1981. Inside The Multinationals — The Economics of International Markets. 1981.

Salis, S. 2008. Foreign Acquisition and Firm Productivity : Evidence from Slovenia. *The World Economy*. 31:1030–1048.

Sambandam, B.R. 2003. Cluster Analysis Gets Complicated. *Journal of Marketing Research*. 15(1):16–21.

Santiago, C. 1987. The impact of foreign direct investment on export structure and employment generation. *World Development*. 15(3):317–328.

Saunders, R.S. 1982. The Determinants of Interindustry Variation of Foreign Ownership in Canadian Manufacturing. *The Canadian Journal of Economics*. 15(1):77–84. [Online], Available: https://www.jstor.org/stable/134670 Accessed: 23-05-2019 01:23 UTC.

Selim, T.H. & Salem, A.S. 2010. Global Cement Industry: Competitive and Institutional Dimensions. *Munich Personal RePEc Archive*. (24464). [Online], Available: https://mpra.ub.uni-muenchen.de/24464/.

Shrivastav, S.M. & Kalsie, A. 2018. The Relationship between Foreign Ownership and Firm Performance in India:An Empirical Analysis. *Artha Vijnana: Journal of The Gokhale Institute of Politics and Economics*. 59(2):152.

Sicom, T. 2017a. Société Internationale de Plantations d ’ Hévéas First Quarter financials 2017. 1–3.

Sicom, T. 2017b. Société Internationale de Plantations d ' Hévéas. 2–3.

Siddharthan, N.S. & Nollen, S. 2007. MNE Affiliation , Firm Size and Exports Revisited : A Study of Information Technology Firms in India. *Journal of Development Studies*. 40(6):146--168.

Simionescu, M. 2016. The relation between economic growth and foreign direct investment during the economic crisis in the European Union. *Zbornik radova Ekonomskog fakulteta u Rijeci*. 34(1):187–213.

Sneath, P.H.A. & Sokal, R.R. 1963. The Principles and Practice of Numerical Taxonomy. *Taxon*. 12(5):190–199.

Stoian, C. & Filippaios, F. 2008. Dunning 's eclectic paradigm : A holistic , yet context specific framework for analysing the determinants of outward FDI Evidence from international Greek investments. *International Business Review*. 17(3):349–367.

Šulc, Z. & Řezanková, H. 2014. Evaluation of Recent Similarity Measures for Categorical Data. *International Scientific Conference*. 249–258.

Sutton, J. 2014. *An enterprise map of Mozambique*. The International Growth Centre, London Publishing Partnership, www.londonpublishingpartnership.co.uk.

Sutton, J. & Kpentey, B. 2012. *An Enterprise Map of Ghana*. The International Growth Centre, London Publishing Partnership, www.londonpublishingpartnership.co.uk.

Sutton, J. & Olomi, D. 2012. *An Enterprise Map of Tanzania*. London: The International Growth Centre, London Publishing Partnership, www.londonpublishingpartnership.co.uk.

Sutton and Langmead. 2013. *An Enterprise Map of Zambia*. UK: International Growth Centre in association with the London Publishing Partnership.

Tambunlertchai, S. 2009. Foreign direct investment and export performance in thailand. Wesleyan University.

Tanner-Smith, E.E. & Lipsey, M.W. 2014. Identifying Baseline Covariates for Use in Propensity Scores: A Novel Approach Illustrated for a Nonrandomized Study of Recovery High Schools. *Peabody Journal of Education*. 89(2):183–196.

Tanton, J., Dodd, L.J., Woodfield, L. & Mabhala, M. 2015. Eating Behaviours of British

University Students : A Cluster Analysis on a Neglected Issue. *Advances in preventive medicine*. 2015(639239):8. [Online], Available: http://dx.doi.org/10.1155/2015/639239.

Timofeev, R. & Hardle, W. 2004. Classification and regression trees (CART) Theory and Applications. Humboldt University, Berlin.

Transparence, L.A. & Les, D. 2015. *Democratic Republic of Congo Executive Committee Report:Extractive Industries Transparence Initiative 2015*.

Tryfos, P. 2002. Cluster analysis. In Vol. 12 *Methods for Business Analysis and Forecasting*. 199–227.

Tsaurai, K. 2017. The Dynamics of Foreign Direct Investment in BRICS Countries Kunofiwa. *Journal of Economics and Behavioral Studies*. 9(3):101–112.

Ullah, I., Shah, M. & Khan, F.U. 2014. Domestic Investment, Foreign Direct Investment, and Economic Growth Nexus: A Case of Pakistan. *Economics Research International*. 2014:1–5.

Ullrich-french, S. & Cox, A. 2009. Using Cluster Analysis to Examine the Combinations of Motivation Regulations of Physical Education Students. *Journal of Sport and Exercise Psychology*. 31(3):358–379.

UNCTAD. 2002. *World Investment Report 2002:Transnational corporations and export competitiveness*. New York and Geneva

UNCTAD. 2014. *World Investment Report 2014: Investing in the SDGs*. Geneva.

UNCTAD. 2017. *World Investment Report 2017*. Geneva. [Online], Available: http://unctad.org/en/PublicationChapters/wir2017ch3_en.pdf.

Vignes, B. 2012. *Société Internationale de Plantations d'Hévéas: Annual Report 2011*.

Vogt, W. & Nagel, D. 1992. Cluster Analysis in Diagnosis. *Journal of Clinical Chemistry*. 38(2):182–198.

Wabiga, P. & Nakijoba, S. 2018. High Technology Exports , Gross Capital Formation and Economic Growth in Uganda : A Vector Auto Regressive Approach. *International Journal of Business and Economics Research.* 7(6):191–202.

Wakyereza, R.K.S. 2017. The Impact of Foreign Direct Investment on Economic Growth , Employment and Poverty Reduction in Uganda. Victoria University.

Waldkirch, A. 2014. *Foreign Firms and Productivity in Developing Countries*. http://citeseerx. ist. psu. edu/viewdoc/download.

Wang, W. 2014. Foreign Multinational Production in Canadian Manufacturing Sector. *Transnational Corporations Review*. 6(1):26–41.

Wang, J. & Wang, X. 2015a. Benefits of foreign ownership : Evidence from foreign direct investment in China. *Journal of International Economics*. 97(2):325–338.

Wang, J. & Wang, X. 2015b. Benefits of foreign ownership: Evidence from foreign direct investment in China. *Journal of International Economics*. 97(2):325–338.

Wang, P., Alba, J.D. & Park, D. 2013. Determinants of Different Modes of FDI : Firm-Level Evidence from Japanese FDI into the US. *Open Economies Review*. 24:425–446.

Weche Geluebcke, J.P. 2015. The impact of foreign takeovers: comparative evidence from foreign and domestic acquisitions in Germany. *Journal of Applied Economics*. 47(8):739–755.

Williamson, O.E. 1975. Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of International Organizations. *New York Free Press*. 2630.

Wilson, N. & Cacho, J. 2007. Linkage Between foreign Direct Investment , Trade and Trade Policy:An Economic Analysis with Application to the Food Sector in OECD Countries and Case Studies in Ghana, Mozambique, Tunisia and Uganda. *OECD Trade Policy Papers*. 50. [Online], Available: http://dx.doi.org/10.1787/152275474424.

Wolf, B.. 1977. Industrial diversification and internalisation: some empirical evidence. *Journal of Industrial Economics*. 26(2):177–191.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed. Massachusetts: The MIT Press Cambridge, Massachusetts London, England.

Yada, B., Tukamuhabwa, P., Kim, D., Skilton, R.A., Alajo, A. & Mwanga, R.O.M. 2010. Characterization of Ugandan Sweetpotato Germplasm Using Fluorescent Labeled Simple Sequence Repeat Markers. *Journal of HortScience*. 45(2):225–230.

Yan, M. 2005. Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion. Virginia Polytechnic Institute and State University.

Yim, O. & Ramdeen, K.T. 2015. Hierarchical Cluster Analysis: Comparison of Three

Linkage Measures and Application to Psychological Data. *The Quantitative Methods for Psychology*. 11(1):8–21.

Yoo, J., Gray, A., Roden, J., M.Fayyad, U., Carvalho, R.R. de & Djorgovski, S.G. 1996. Analysis of Digital POSS-II Catalogs Using Hierachical Unsupervised Learning Algorithms. In Vol. 101. Astronomical Society of the Pacific *Astronomical Data Analysis Software and Systems V*. 41.

Yurtseven, A.E. & Tandoğan, S. 2012. Patterns of Innovation and Intra-industry Heterogeneity in Turkey. *International Review of Applied Economics*. 26(5):657–671. [Online], Available: http://www.stps.metu.edu.tr.

Zambia Government. 2006. *"A prosperous Middle-income Nation By 2030"*.

Zebregs, M.. 1998. *Can the neoclassical model explain the distribution of foreign direct investment across developing countries?* The International Monetary Fund.

Zhang, Z., Kim, H.J., Lonjon, G. & Zhu, Y. 2019. Balance diagnostics after propensity score matching. *Annals of translational medicine*. 7(I):1–8.

Zubin, J. 1938. A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology*. 33(4):508.

Appendices

Appendix 1

Table 2 A: P-values for t-test statistics for clusters of all the datasets

Clusters	ln_emp	lb_prdvty	capital_int	mat/worker	work_educ	mgt_exp	wage	exports
NData 2006								
1~2	0.000	0.000	0.005	0.000	-	0.003	0.000	0.001
NData2013								
1~2	0.000	0.000	0.123	0.019	0.670	0.002	0.182	0.001
1~3	0.794	0.143	0.233	0.523	<0.000	0.525	0.506	0.374
2~3	0.000	0.001	0.758	0.152	<0.000	0.066	0.129	0.089
CNData								
1~2	0.000	0.000	0.009	0.000	<0.000	0.000	0.000	0.015
1~3	0.000	0.000	0.002	0.000	0.824	0.000	0.000	0.000
2~3	0.448	0.363	0.938	0.389	<0.000	0.365	0.682	0.538

Source: Author's own computations based on World Bank Enterprise Survey Data, 2006/13 panel

Table 2 B: Mean actual employment, output and firm exports for clusters of CNData

	Average employment	Average Output (000'000/=)	Average Exports (000'000/=)
Cluster1	41.77	8397.7	3982.3 (47.4)
Cluster2	15.66	1288.7	506.8 (39.3)
Cluster3	22.21	428.9	233.7 (54.5)

Source: Author's own computation based on World Bank Enterprise Survey data 06/13

Table 2 C: p-values for the correlation matrix variables used in firm clustering

Variables	empt	fdi_stake	prdvty	wage	mgt exp	k in'ty	mat/w'k	w_educ	exports
empt									
fdi_stake	0.00								
prdvty	0.00	0.00							
wage	0.23	0.00	0.00						
mgt exp	0.00	0.00	0.17	0.56					
k in'ty	0.00	0.00	0.00	0.00	0.01				
mat/w'k	0.02	0.00	0.00	0.00	0.02	0.00			
w_educ	0.00	0.04	0.60	0. 50	0.13	0.33	0.90		
exports	0.30	0.57	0.91	0.34	0.04	0.21	0.02	0.00	

Source: Author's own computation based on World Bank Enterprise Survey data 06/13

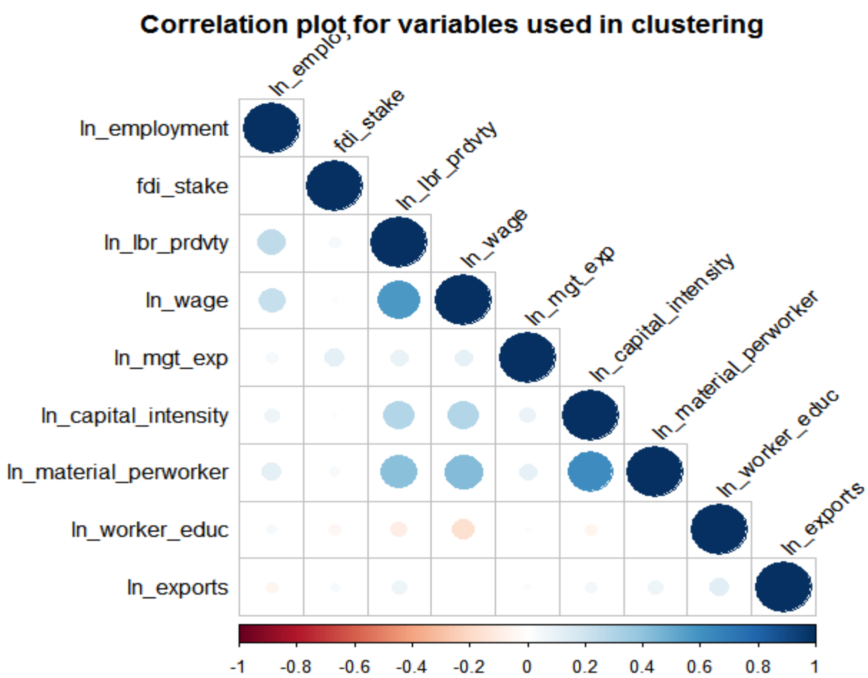


Figure 2 A: Correlation plot for variables used in clustering

Table 2 D: Regression results using clusters as regressands on selected variables

	Dependent Variables				
Independent Variables	emp't	lbr_prdvty	mgt_exp	wage	work_educ
as.factor (Cluster) 1	-	-	-	-	-
as.factor (Cluster) 2	-0.826***	-1.314***	-0.238***	-0.560***	-0.113***
	(0.057)	(0.114)	(0.044)	(0.086)	(0.039)
as.factor (Cluster) 3	-0.815***	-1.250***	-0.127**	-0.593***	-1.934***
	(0.081)	(0.162)	(0.064)	(0.133)	(0.073)
Constant	3.0689***	16.767***	2.416***	14.185***	2.636***
	(0.053)	(0.106)	(0.041)	(0.079)	(0.037)
RSE	1.672	3.014	1.297	2.417	0.37
	(df=1300)	(df=1054)	(df=1282)	(df=1003)	(df=286)

Note: *p<0.1; **p<0.05; ***p<0.01

Source: Author's results generated based on World Bank Enterprise Survey data 0/13

Table 2 E: Principal components for variable clusters of 2013 and overall data

	2006			2013			overall data		
	Cluster1	Cluster2	Cluster3	Cluster1	Cluster2	Cluster3	Cluster1	Cluster2	Cluster3
PC1	1.326	2.588	1.000	1.319	2.463	2.019	1.319	2.735	1.463
PC2	0.674	0.372	-	1.146	0.365	0.595	1.146	0.740	0.537
PC3	-	0.040	-	0.535	0.172	0.385	0.535	0.354	-

Source: Author's own computations based on World Bank Enterprise Survey data 06/13

Table 2 F: Regression selected variables, firm ownership and other controls

Dependent variable:	ln_wage			ln_mgt_exp		
Independent Variables	M1	M2	M3	M1	M2	M3
fdi_dummy	0.54***		0.18***	0.23***		0.21***
	(0.040)		(0.059)	(0.019)		(0.023)
fdi_dummy_cat 10-40%		0.88***			0.40***	
		(0.194)			(0.102)	
fdi_dummy_cat 40-60%		1.17***			0.40***	
		(0.141)			(0.063)	
fdi_dummy_cat 60-99%		1.38***			0.55***	
		(0.102)			(0.052)	
fdi_dummy_cat all_foreign		0.34***			0.130***	
		(0.044)			(0.022)	
sector_type_retail			0.39***			-0.40***
			(0.056)			(0.019)
sector_type_others			0.71***			-0.25***
			(0.055)			(0.018)
location_catkampala			0.23**			-0.39***
			(0.091)			(0.031)
location_catlira			0.50***			-0.39***
			(0.186)			(0.056)
location_catmbale			-0.143			-0.74***
			(0.143)			(0.044)
location_catmbarara			0.55***			-0.87***
			(0.134)			(0.047)
location_catwakiso			0.141			-0.51***
			(0.114)			(0.038)
Observations	1002	1001	441	1281	1281	723
RSE	1.170	1.117	1.334	0.785	0.731	0.643

Note: **p<0.1; **p<0.05; ***p<0.01, figures in parenthesis are standard errors*

Source: Author's own results generated based on World Bank Survey Data 2006/13 panel

Table 2 G: Descriptive Statistics, Weighted Mean and Median (real) values NData 2006

Variable	Cluster1 N = 89	Cluster 2 N = 474
Mean		
ln_employment	3.22	2.53
ln_lbr_prdvty	16.83	15.74
fdi_stake	93.42	0.24
ln_capital_intensity	14.77	14.27
ln_material_perworker	16.08	15.04
ln_mgt_exp	2.21	2.02
ln_wage	14.64	13.89
ln_worker_educ	-	-
ln_exports_absolute	20.00	18.08
actual employment	51.01	27.55
Median		
ln_employment	3.22	2.30
ln_lbr_prdvty	16.80	15.59
fdi_stake	100.0	0.00
ln_capital_intensity	15.09	14.31
ln_material_perworker	16.01	14.97
ln_mgt_exp	2.30	2.08
ln_wage	14.27	13.92
ln_worker_educ	-	-
ln_exports_absolute	19.95	17.67
actual employment	25.00	10.00

Source: Author's own computations based on World Bank Enterprise Survey Data, 2006/13 panel

Table 2 H: Descriptive statistics, weighted mean and median (real) values of NData 2013

Variable	Cluster1 N = 93	Cluster 2 N = 102	Cluster 3 N = 567
Mean			
ln_employment	2.22	2.95	2.09
ln_lbr_prdvty	15.21	17.62	15.58
fdi_stake	0.19	94.23	0.13
ln_capital_intensity	13.13	15.66	14.98
ln_material_perworker	12.66	14.05	13.89
ln_mgt_exp	2.21	2.45	2.04
ln_wage	13.10	13.30	12.53
ln_worker_educ	2.51	2.45	0.78
ln_exports_absolute	18.07	22.01	18.56
actual employment	18.26	34.50	14.99
Median			
ln_employment	2.08	2.83	2.08
ln_lbr_prdvty	15.03	16.62	15.23
fdi_stake	0.00	100.0	0.00
ln_capital_intensity	15.24	15.32	15.05
ln_material_perworker	12.63	14.23	13.76
ln_mgt_exp	2.30	2.40	1.95
ln_wage	13.22	13.10	13.13
ln_worker_educ	2.56	2.64	0.69
ln_exports_absolute	17.96	22.98	19.06
actual employment	8.00	17.00	8.00

Source: Author's own computations based on World Bank Enterprise Survey Data, 2006/13 panel

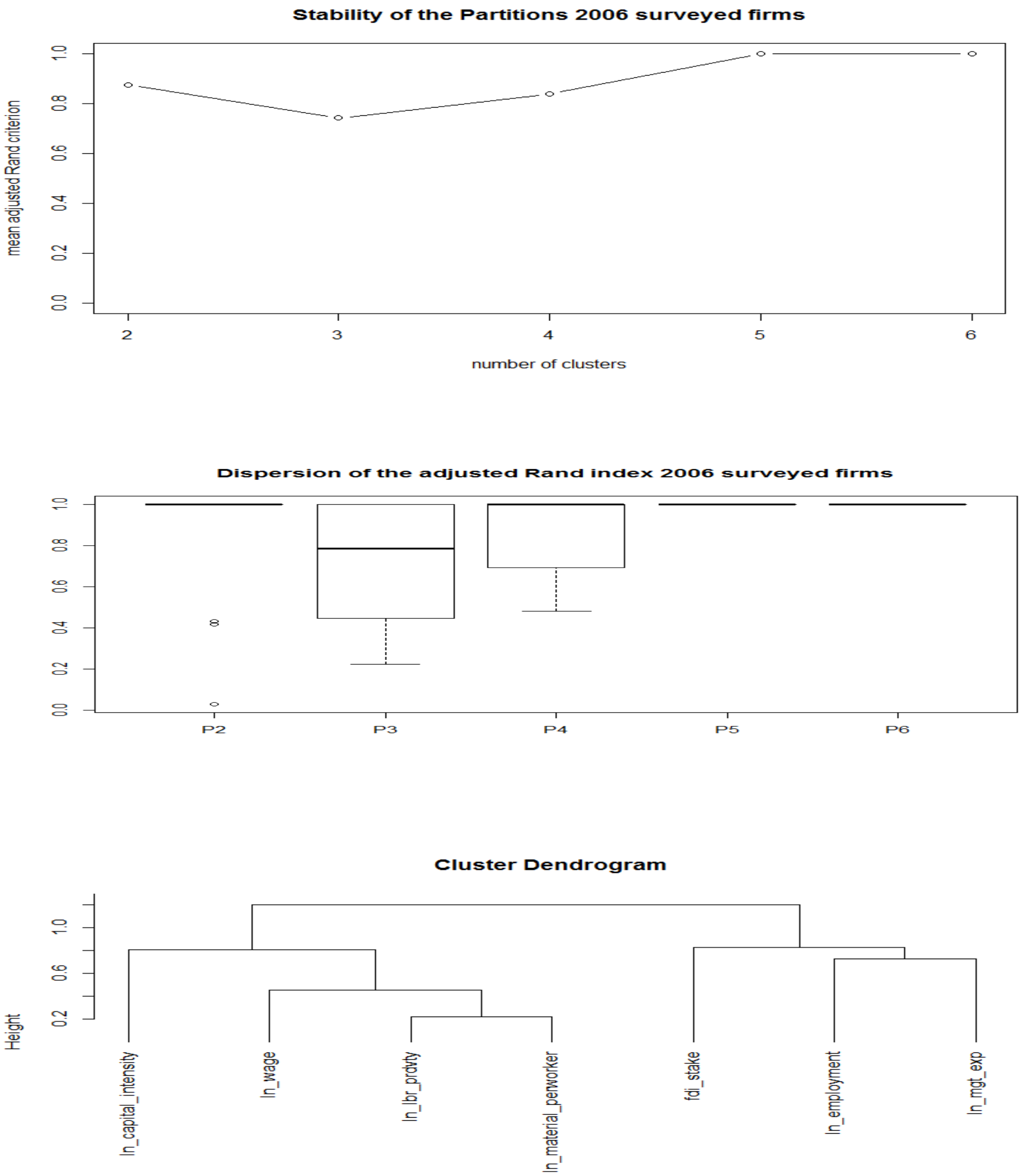


Figure 2 B: Dendrogram, stability plot and box plot for 2006 variable clusters

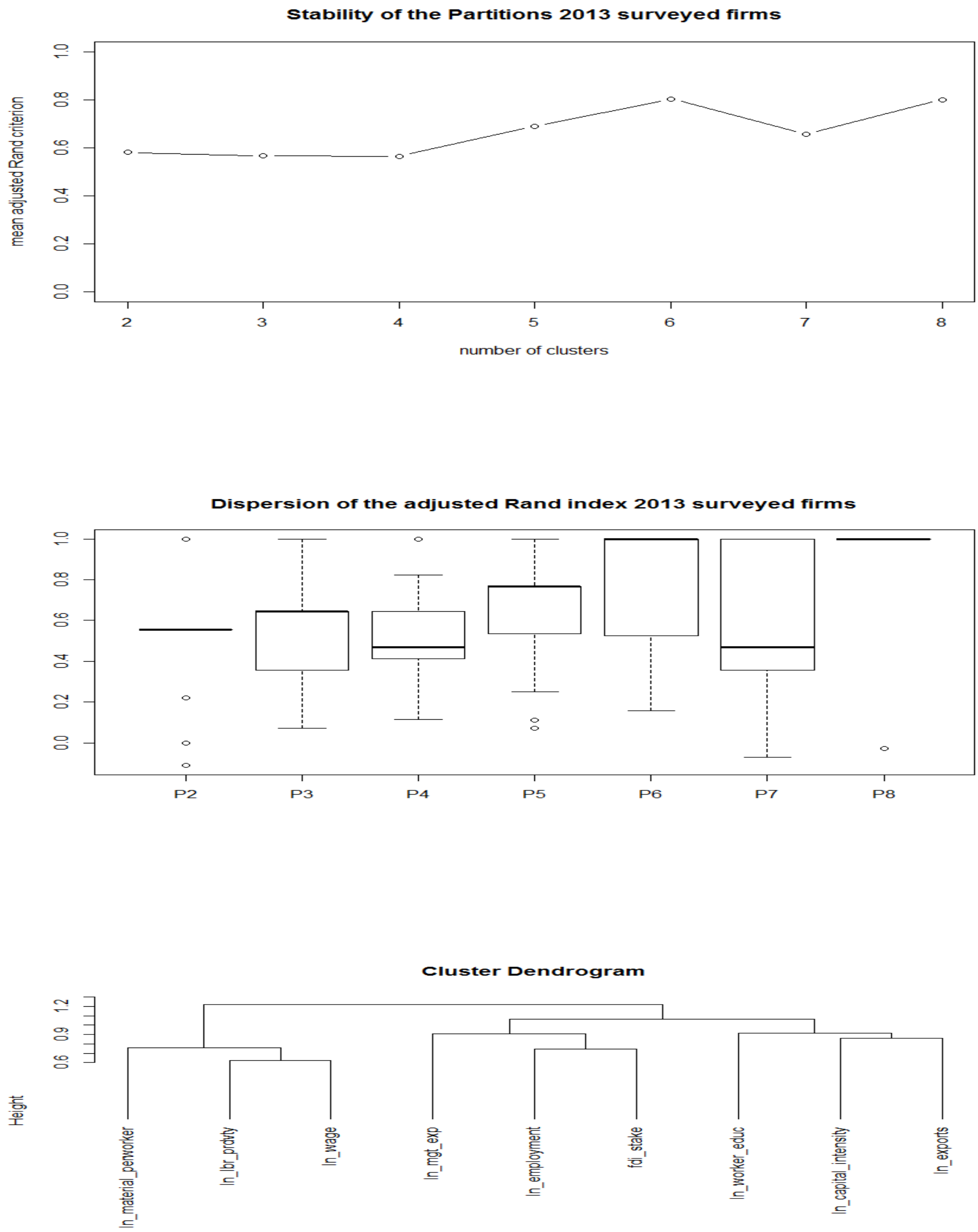


Figure 2 C: Stability plot, box plot and dendrogram for 2013 variable clusters

Appendix 2

Table 3A 1: Descriptive Statistics, mean and median (real) values SSA

Variable	*All* (N=6497)	*Foreign* (N=2313)	*Domestic* (N=4184)	*p-values*
Mean				
Lnwage	2.67	2.77	2.62	0.004
fdi-stake	30.6	85.97	0.00	0.000
lnemployment	4.03	4.39	3.83	0.000
lncapital-intensity	3.24	3.37	3.17	0.007
Lnsales	7.83	8.60	7.37	0.000
lnlbr_prdvty	4.02	4.18	3.92	0.002
lnmaterial_perworker	3.61	3.69	3.56	0.038
lnemployee_training	3.49	3.63	3.40	0.027
Lnexports	7.55	7.87	7.11	0.000
lnresearch_invest	4.15	5.56	4.14	---
Median				
Lnwage	2.61	2.87	2.46	
fdi-stake	0.00	100	0.00	
lnemployment	3.91	4.32	3.69	
lncapital-intensity	3.07	3.17	3.00	
Lnsales	7.75	8.62	7.36	
lnlbr_prdvty	4.01	4.23	3.89	
lnmaterial_perworker	3.44	3.57	3.36	
lnemployee_training	3.20	3.29	3.12	
Lnexports	7.44	7.63	7.18	
lnresearch_invest	3.81	5.56	3.81	

Source: Author's own results based on AIS Survey Data, 2010

Table 3A 2: Descriptive Statistics, mean and median (real) values for income groups

Variable	Uganda		Low income Ug excl		Lower Middle Income	
	F	D	F	D	F	D
Mean						
lnwage	3.05	2.63	4.58	4.29	6.53	5.90
lnemployment	4.18	3.59	4.46	3.96	4.37	3.55
lncapital_intensity	3.11	2.65	**5.15**	**5.01**	**6.31**	**5.92**
lnsales	9.39	7.69	11.6	10.3	13.3	11.3
lnlbr_prdvty	4.99	4.08	6.83	6.26	8.37	7.50
lnmaterial_perworker	4.21	3.32	**5.99**	**5.53**	7.52	6.93
lnemployee_training	4.50	3.48	5.79	5.38	7.39	6.70
lnexports	8.45	8.14	11.0	10.0	12.5	10.8
lnresearch_invest	8.52	3.55	-	5.46	-	7.17
actual employment	199	115	272	159	261	110
Median						
lnwage	2.95	2.63	4.66	4.11	6.54	5.86
lnemployment	4.09	3.40	4.38	3.81	4.27	3.47
lncapital_intensity	3.08	2.60	4.93	4.87	6.43	6.03
lnsales	9.22	7.30	11.8	10.2	13.3	11.2
lnlbr_prdvty	4.94	3.96	7.05	6.25	8.43	7.43
lnmaterial_perworker	4.16	3.22	5.89	5.29	7.46	6.93
lnemployee_training	4.16	3.48	5.72	5.19	7.36	6.59
lnexports	8.47	8.21	11.2	10.2	12.5	10.5
lnresearch_invest	8.52	3.22	-	5.10	-	6.94
actual employment	60.0	30.0	80	45	72	32

Source: Author's own computations based on AIS Survey Data, 2010

Table 3 A

Table 3 B: Absolute composition of firms by Income groupings and sector

Clusters	Uganda	LY	LM	F	D	Manuf	Serv	Cluster Total
Cluster 1	412	2772	1000	2	4182	2335	1849	4184
Cluster2	29	227	57	313	0	185	128	313
Cluster3	340	946	266	1552	0	875	677	1552
Cluster4	32	300	112	444	0	296	148	444
Cluster5	0	4	0	2	2	1	3	4
Total	**813**	**4249**	**1435**	**2313**	**4184**	**3692**	**2805**	**6497**

Source: Author's own computations based on AIS Survey Data, 2010

Table 3 C: *p* -values for t-test statistics for the five clusters

	Variables										
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk	inInv	new_inv
1~ 2	0.000	<0.000	0.001	0.000	0.003	0.001	0.002	0.001	<0.000	<0.000	0.000
1~ 3	0.657	<0.000	0.372	0.000	0.610	0.186	0.521	0.006	<0.000	<0.000	0.643
1~ 4	0.000	<0.000	0.000	<0.000	0.000	0.000	0.001	0.000	<0.000	<0.000	0.000
1~ 5	0.410	0.698	0.951	-	-	0.955	-	-	0.185	0.942	0.159
2~ 3	0.000	0.000	0.001	0.007	0.002	0.000	0.001	0.094	<0.000	0.006	0.000
2~ 4	0.825	0.728	0.444	0.411	0.355	0.796	0.993	0.556	<0.000	0.610	0.367
2~ 5	0.235	0.508	0.698	-	-	0.696	-	0.235	0.281	0.427	0.328
3~ 4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	<0.000	0.011	0.000
3~ 5	0.425	0.789	0.989	-	-	0.894	-	-	0.133	0.447	0.165
4~ 5	0.226	0.481	0.634	-	-	0.670	-	-	0.897	0.431	0.394

Source: Author's own computations based on AIS Survey Data, 2010

Table 3 D: Hopkins statistic tests and optimal clusters for the data sets

Data set	Optimal Clusters	Value index	**H** statistic
SSA	5.0	53394.01	0.003
Ethiopia	5.0	1690.22	0.024
Kenya	3.0	981.73	0.019
Nigeria	6.0	1893.91	0.014
Tanzania	5.0	2198.86	0.015
Low income	3.0	32155.99	0.002
Lower middle-income	3.0	4688.98	0.009

Source: Author's own computations based on AIS Survey Data, 2010

Table 3 E: *p*-values for Correlation Matrix for SSA Data set

Variables	1	2	3	4	5	6	7	8	9	10
1. employment	-	0.73	0.75	0.84	0.83	0.00	0.18	0.00	0.00	0.00
2. lbr_prdvty		-	0.00	0.00	0.00	0.00	0.90	0.34	0.00	0.15
3. capital_intensity			-	0.00	0.74	0.00	0.89	0.47	0.00	0.04
4. material_perworker				-	0.00	0.00	0.90	0.73	0.00	0.03
5. wage					-	0.42	0.85	0.31	0.17	0.12
6. exports						-	0.79	0.00	0.00	0.00
7. research_investment							-	0.04	0.65	0.88
8. employee_training								-	0.00	0.75
9. sales									-	0.07
10.fdi_stake										-

Source: Author's own computations based on AIS Survey Data, 2010

Table 3F 1: Real mean values for clusters of Ethiopian firms

	Variables											
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	Total
Mean												
Cluster 1	4.14	0.00	4.51	5.20	11.1	6.40	5.90	5.76	10.4	6.84	215	**438**
Cluster 2	5.06	33.3	4.41	11.9	11.8	7.36	6.95	-	7.81	9.53	170	**3**
Cluster 3	3.73	0.00	4.14	3.96	17.9	13.0	4.26	-	-	-	81	**2**
Cluster 4	4.51	99.2	4.62	6.22	11.5	6.62	6.08	7.20	11.3	-	271	**91**
Cluster 5	4.51	50.7	4.81	5.97	11.5	6.61	6.26	6.33	10.9	-	198	**35**
Ethiopia	4.22	19.2	4.55	5.43	11.2	6.47	5.95	6.08	10.6	6.94	222	**569**

Source: Author's own computations based on AIS Survey Data, 2010

Table 3F 2: *p*-values for t-test statistics for the Ethiopian five clusters

	Variables							
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt
1~ 2	0.000	0.934	-	0.758	0.319	0.261	-	-
1~ 3	0.502	0.701	-	-	-	0.156	-	-
1~ 4	0.019	0.516	0.000	0.222	0.274	0.000	0.001	0.017
1~ 5	0.058	0.095	0.009	0.312	0.478	0.220	0.319	0.429
2~ 3	0.186	0.852	-	-	-	0.044	-	-
2~ 4	0.001	0.869	-	0.879	0.419	0.325	-	-
2~ 5	0.007	0.511	-	0.911	0.420	0.419	-	-
3~ 4	0.284	0.000	-	-	-	0.115	-	-
3~ 5	0.273	0.524	-	-	-	0.064	-	-
4~ 5	0.996	0.382	0.459	0.858	0.984	0.574	0.171	0.503

Source: Author's own computations based on AIS Survey Data, 2010

Table 3G 1: Real mean values for clusters of Kenyan firms

Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	Total
Mean												
Cluster 1	3.58	1.86	4.41	3.73	10.5	5.75	4.87	4.80	9.12	6.07	234	**352**
Cluster 2	3.93	43.0	5.26	5.16	11.6	6.36	5.73	6.17	11.0	9.53	556	**75**
Cluster 3	3.93	92.7	4.85	4.75	11.4	6.31	5.77	5.55	10.8	-	386	**187**
Kenya	3.73	34.5	4.64	4.25	11.0	6.06	5.26	5.11	10.1	6.07	320	**614**

Source: Author's own computations based on AIS Survey Data, 2010

Table 3G 2: *p*-values for t-test statistics for the Kenyan clusters

Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 2	0.006	0.000	0.000	0.000	0.000	0.000	0.004	0.000	<0.000
1~ 3	0.001	0.000	0.000	0.000	0.000	0.000	0.010	0.000	<0.000
2~ 3	0.958	0.037	0.128	0.411	0.770	0.847	0.192	0.516	<0.000

Source: Author's own computations based on AIS Survey Data, 2010

Table 3H 1: Real mean values for clusters of Nigerian firms

Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	**Total**
Mean												
Cluster 1	2.42	0.68	4.16	3.06	8.44	4.33	3.66	4.16	8.94	4.33	177	**482**
Cluster 2	3.28	24.3	5.69	4.37	14.9	8.60	5.00	6.42	8.42	-	549	**6**
Cluster 3	2.59	40.0	5.46	4.05	10.8	5.44	5.17	4.38	-	10.2	337	**5**
Cluster 4	2.62	99.4	4.61	3.32	8.84	4.60	4.03	4.40	9.37	-	251	**56**
Cluster 5	3.11	59.7	5.17	3.94	10.2	5.18	4.07	4.46	7.44	-	427	**46**
Cluster 6	9.41	100	4.53	2.50	-	-	14.0	7.23	-	-	93	**1**

Nigeria	2.52	15.3	4.31	3.18	8.73	4.49	3.77	4.26	8.73	4.48	208	**596**

Source: Author's own computations based on AIS Survey Data, 2010

Table 3H 2: *p*-values for t-test statistics for the key clusters in Nigeria

	Variables								
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 2	0.294	0.064	0.107	0.000	0.000	0.307	0.034	-	0.130
1~ 3	0.424	0.053	0.583	0.116	0.170	0.004	0.485	-	0.184
1~ 4	0.362	0.022	0.425	0.335	0.411	0.188	0.528	0.684	<0.000
1~ 5	0.001	0.000	0.017	0.000	0.006	0.231	0.461	0.217	<0.000
2~ 3	0.396	0.776	0.855	0.032	0.008	0.893	0.035	-	0.592
2~ 4	0.419	0.160	0.186	0.000	0.000	0.455	0.025	-	0.002
2~ 5	0.822	0.475	0.568	0.000	0.000	0.479	0.026	-	0.042
3~ 4	0.909	0.154	0.683	0.166	0.280	0.010	0.965	-	0.072
3~ 5	0.067	0.609	0.953	0.635	0.720	0.019	0.852	-	0.466
4~ 5	0.085	0.051	0.177	0.008	0.162	0.918	0.898	0.174	<0.000

Source: Author's own computations based on AIS Survey Data, 2010

Table 3I 1: Real mean values for clusters of Tanzanian firms

	Variables											
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	**Total**
Mean												
Cluster 1	0.80	0.00	3.62	1.25	5.40	1.94	1.63	1.86	5.55	2.54	73	**304**
Cluster 2	4.17	0.00	4.70	1.68	9.53	5.14	4.42	6.46	6.04	6.28	115	**2**
Cluster 3	0.65	0.00	3.95	-	14.3	10.4	1.31	-	13.5	-	52	**1**
Cluster 4	0.81	98.5	4.32	1.51	7.30	2.79	2.15	2.40	6.43	-	137	**112**
Cluster 5	1.17	52.4	4.78	1.80	7.96	3.09	2.67	2.57	6.40	-	407	**40**
Tanzania	0.87	28.6	3.89	1.39	6.12	2.29	1.88	2.19	6.07	2.65	117	**459**

Source: Author's own results based on AIS Survey Data, 2010

Table 3I 2: *p*-values of t-test statistics for the three key clusters in Tanzania

	Variables								
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 4	0.936	0.000	0.165	0.000	0.000	0.001	0.073	0.085	<0.000
1~ 5	0.039	0.000	0.095	0.000	0.000	0.001	0.044	0.037	<0.000
4~ 5	0.048	0.037	0.398	0.142	0.342	0.087	0.674	0.973	<0.000

Source: Author's own results based on AIS Survey Data, 2010

Table 3J 1: Real mean values for clusters of Lower income economies

	Variables											
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	Total
Mean												
Cluster 1	2.38	0.06	3.91	3.10	7.12	3.70	3.32	3.22	7.05	4.09	156	**3189**
Cluster 2	2.26	98.0	4.33	3.01	7.86	3.55	3.14	3.09	7.24	-	243	**1470**
Cluster 3	3.02	47.0	4.63	4.17	9.40	4.91	4.39	4.02	8.51	-	278	**403**
Overall	2.41	32.2	4.09	3.18	7.55	3.77	3.35	3.27	7.32	4.10	191	**5062**

Source: Author's own results based on AIS Survey Data, 2010

Table 3J 2: *p*-values of t-test statistics for clusters of lower-income economies

	Variables								
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 2	0.075	<0.000	0.403	0.000	0.159	0.045	0.365	0.302	<0.000
1~ 3	0.000	<0.000	0.000	<0.000	0.000	0.000	0.001	0.000	<0.000
2~ 3	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	<0.000

Source: Author's own results based on AIS Survey Data, 2010

Table 3K 1: Real mean values for clusters of Lower middle-income economies

	Variables											
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnrsch*	*empt*	Total
Mean												
Cluster1	3.16	1.10	3.61	3.33	8.39	4.65	4.14	3.93	7.93	4.19	116	**1046**
Cluster2	3.81	99.2	4.24	3.40	10.4	5.58	4.84	4.51	10.0	-	257	**283**
Cluster3	3.48	60.7	4.67	3.86	10.2	5.28	4.58	4.27	8.39	-	301	**106**
Overall	3.32	24.9	3.81	3.39	8.93	4.88	4.31	4.07	8.86	4.28	158	**1435**

Source: Author's own results based on AIS Survey Data, 2010

Table 3K 2: *p*-values of t-test statistics for clusters of lower middle-income economies

	Variables								
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 2	0.000	0.000	0.591	0.000	0.000	0.000	0.009	0.000	<0.000
1~ 3	0.042	0.000	0.032	0.000	0.009	0.047	0.301	0.389	<0.000
2~ 3	0.069	0.014	0.093	0.582	0.271	0.303	0.502	0.006	<0.000

Source: Author's own results based on AIS Survey Data, 2010

Table 3L 1: Mean values for SSA clusters without fdi_stake as a clustering variable

	Variables									
Cluster	*lnw*	*fdistk*	*lnempt*	*lncapint*	*lnsales*	*lnlbrpvty*	*lnmat*	*lntrn*	*lnexp*	*lnR&D*
Mean										
Cluster 1	2.61	29.0	3.76	3.16	7.40	3.90	3.53	3.02	6.89	3.45
Cluster 2	3.04	41.3	6.47	3.59	10.6	4.40	3.98	5.38	9.01	5.47
Cluster 3	2.53	44.3	8.42	2.93	10.9	3.15	2.42	4.09	9.77	3.26
Cluster 4	4.57	67.1	5.92	4.99	13.8	7.80	7.09	6.03	13.9	-
SSA	2.67	30.6	4.03	3.24	7.83	4.02	3.61	3.49	7.55	-

Source: Author's own results based on AIS Survey Data, 2010

Table 3L 2: *p*-values for t-test statistics for the fdi-less clusters

	Variables								
Clusters	wage	emp't	cap'int	sales	prdvty	material	training	expt	fdi_stk
1~ 2	0.006	<0.000	0.000	<0.000	0.000	0.000	<0.000	<0.000	0.000
1~ 3	0.868	<0.000	0.620	0.000	0.144	0.030	0.112	0.000	0.044
1~ 4	0.000	0.000	0.000	<0.000	<0.000	0.000	0.000	<0.000	0.000
2~ 3	0.273	<0.000	0.181	0.541	0.023	0.004	0.064	0.218	0.680
2~ 4	0.000	0.015	0.000	0.000	<0.000	0.000	0.090	<0.000	0.000
3~ 4	0.000	0.000	0.001	0.000	0.000	0.000	0.013	0.000	0.018

Source: Author's own computations based on AIS Survey Data, 2010

Appendix 3

Table 4 A: Sample composition

Sub-sector	*Frequency*	*Percentage*	*cum-frequency*
Bakery	180	10.9	10.9
Garment	360	21.9	32.8
Textile	36	2.2	35.0
Wood	72	4.4	39.4
Furniture	312	19.0	58.4
Metal	324	19.7	78.1
Machines	48	2.9	83.9
Chemical	48	2.9	83.9
SSRII	12	0.7	84.7
Food	216	13.1	97.8
Drink	36	2.2	100
Total	**1644**	**100**	---
Location			
Accra	876	52.9	52.9
Cape coast	36	2.2	55.1
Kumasi	624	37.7	92.8
Takoradi	120	7.2	100
Total	**1656**	**100**	---

*Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*

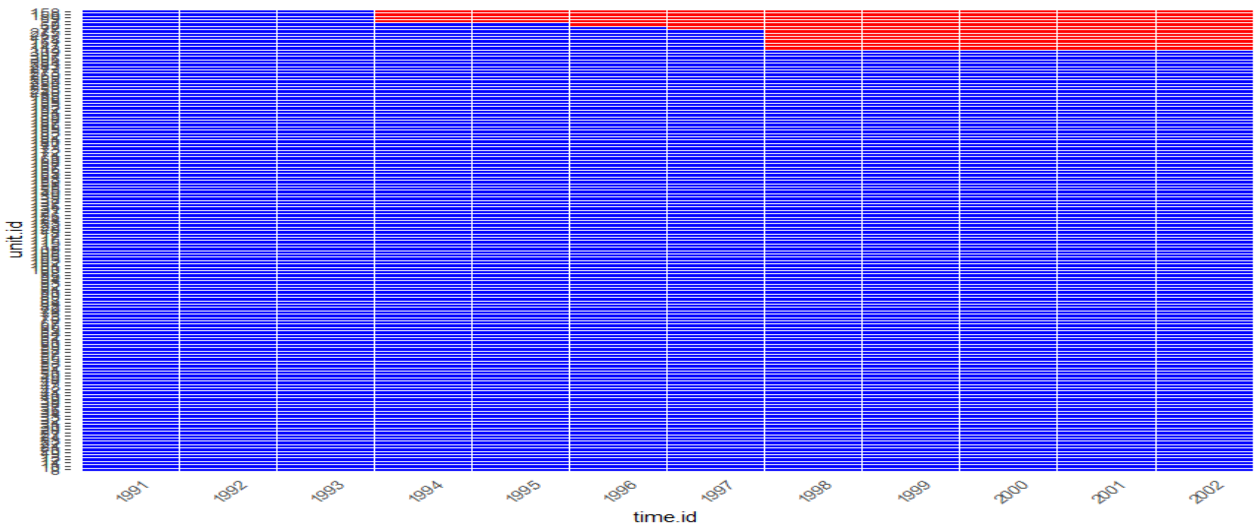


Figure 4 A: Treatment variation plot for overall sample

Table 4 B: Matched sets before refinement

Firm ID	Year	Matched set size
7	1994	134
68	1996	133
69	1994	134
77	1997	132
113	1998	126
147	1998	126
150	1994	134
178	1998	126
158	1994	134
267	1998	126
275	1998	126
161	1998	126

Note: Generated using R package "Panelmatch"

Table 4 C: Weighted difference-in-differences estimates of ATT with ps.weight & ps.match

Variables	*F = 0*	*F = 1*	*F = 2*	*F = 3*
Employment	0.015	0.051	0.095	0.037
	(0.094)	(0.072)	(0.110)	(0.115)
Wage	0.222**	0.228**	0.179	0.021
	(0.106)	(0.098)	(0.203)	(0.238)
Productivity	0.159	-0.002	-0.032	-0.089
	(0.416)	(0.418)	(0.324)	(0.288)
Capital	0.043	0.160*	0.180*	0.161
	(0.058)	(0.100)	(0.116)	(0.126)
Output	0.132	0.028	0.002	0.110
	(0.328)	(0.336)	(0.325)	(0.304)
K_int	0.009	-0.050	-0.020	0.075
	(0.188)	(0.151)	(0.209)	(0.231)
Using ps.match				
Employment	0.040	0.005	0.066	0.063
	(0.137)	(0.091)	(0.126)	(0.149)
Wage	0.334***	0.343***	0.465**	0.285
	(0.122)	(0.133)	(0.216)	(0.189)
Productivity	0.148	0.037	-0.034	-0.019
	(0.406)	(0.404)	(0.344)	(0.314)
Capital	0.061	0.187*	0.201*	0.127
	(0.055)	(0.109)	(0.119)	(0.132)
Output	0.093	0.025	-0.024	0.009
	(0.341)	(0.339)	(0.322)	(0.301)
K_int	0.125	0.112	0.208	0.220
	(0.206)	(0.180)	(0.180)	(0.215)

*Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*

Table 4 D: Effects of Foreign Acquisition on firm performance - Estimates with lagged dependent variables

FE-full sample	1	2	3	4	5	6
Variables	employment	wage	productivity	output	capital	skill-int
Fdi	0.106**	0.092*	0.194	0.266*	0.015	0.398*
	(0.046)	(0.054)	(0.158)	(0.142)	(0.086)	(0.235)
lag dep.var	0.480***	0.344***	0.395***	0.429***	0.829***	-.090
	(0.060)	(0.034)	(0.033)	(0.037)	(0.059)	(0.144)
constant	1.383***	6.709***	8.384***	9.473***	2.704***	16.82***
	(0.177)	(0.360)	(0.457)	(0.623)	(0.918)	(0.091)
observations	1194	1398	1156	1157	1217	1062
№ - firms (N)	138	138	138	138	138	137
sector – dum's	Y	Y	Y	Y	Y	Y
year - dum's	Y	Y	Y	Y	Y	Y
R^2- within	0.287	0.365	0.184	0.231	0.642	0.040

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Own results based on sample data